

EXHIBIT M

REQUESTS FOR ADMISSION

ADMISSION NO. 1:

The disparity between the pass rate of black candidates on Written Exam 7029 and the pass rate of white candidates on Written Exam 7029 is equivalent to more than three (3) units of standard deviation.

OBJECTION AND RESPONSE TO ADMISSION NO. 1:

Admits that in comparing the proportion of those White candidates who passed Written Exam 7029 with the proportion of Black candidates who passed Written Exam 7029, there was a difference greater than three units of standard error. As explained in the appendix to defendant's expert report when comparing two groups, the appropriate measure of difference in ranking would be designated by "units of standard error," which although being a subset of "standard deviation" is not synonymous with the term "standard deviation."

ADMISSION NO. 2:

The disparity between the pass rate of Hispanic candidates on Written Exam 7029 and the pass rate of white candidates on Written Exam 7029 is equivalent to more than three (3) units of standard deviation.

OBJECTION AND RESPONSE TO ADMISSION NO. 2:

Admits that in comparing the proportion of those White candidates who passed Written Exam 7029 with the proportion of Hispanic candidates who passed Written Exam 7029, there was a difference greater than three units of standard error. As explained in the appendix to defendant's expert report when comparing two groups, the appropriate measure of difference in ranking would be designated by "units of standard error," which although being a subset of "standard deviation" is not synonymous with the term "standard deviation."

ADMISSION NO. 3:

The disparity between the pass rate of black candidates on Written Exam 2043 and the pass rate of white candidates on Written Exam 2043 is equivalent to more than three (3) units of standard deviation.

OBJECTION AND RESPONSE TO ADMISSION NO. 3:

Admits that in comparing the proportion of those White candidates who passed Written Exam 2043 with the proportion pf Black candidates who passed Written Exam 2043, there was a difference greater than three units of standard error. As explained in the appendix to defendant's expert report when comparing two groups, the appropriate measure of difference in ranking would be designated by "units of standard error," which although being a subset of "standard deviation" is not synonymous with the term "standard deviation."

ADMISSION NO. 4:

The disparity between the pass rate of Hispanic candidates on Written Exam 2043 and the pass rate of white candidates on Written Exam 2043 is equivalent to more than three (3) units of standard deviation.

OBJECTION AND RESPONSE TO ADMISSION NO. 4:

Admits that in comparing the proportion of those White candidates who passed written Exam 2043 with the proportion of Hispanic candidates who passed written Exam 2043, there was a difference greater than three units of standard error. As explained in the appendix to defendant's expert report when comparing two groups, the appropriate measure of difference in ranking would be designated by "units of standard error," which although being a subset of "standard deviation" is not synonymous with the term "standard deviation."

ADMISSION NO. 5:

As a group, black candidates on the eligibility list resulting from Exam 7029 were ranked statistically significantly lower (i.e., had statistically significantly higher list numbers) than white candidates on the eligibility list. For purposes of this Request, "statistically significant" means that there is a 5% or lower likelihood that so large a disparity would occur by chance.

OBJECTION AND RESPONSE TO ADMISSION NO. 5:

Admits that a statistically significant difference exists between the ranked positions of Black candidates, as a group, and the ranked positions of White candidates, as a group, on the eligibility list relating to Exam 7029 but defendant does not admit that any such difference was caused by the Exam.

ADMISSION NO. 6:

The disparity between the ranks of black candidates on the eligibility list resulting from Exam 7029 and the ranks of white candidates on the eligibility list is equivalent to more than three (3) units of standard deviation.

OBJECTION AND RESPONSE TO ADMISSION NO. 6:

Admits that the difference between the ranks of Black candidates on the eligibility list relating to Exam 7029 and the ranks of White candidates on the eligibility list is equivalent to more than three (3) units of standard error so long as there are no controls for sample size, however when controls for sample size, such as those used by plaintiff's expert, are used in analyzing the rankings of these groups then the difference is less than three units of standard error. As explained in the appendix to defendant's expert report when comparing two groups, the appropriate measure of difference in ranking would be designated by "units of standard

error,” which although being a subset of “standard deviation” is not synonymous with the term “standard deviation.”

ADMISSION NO. 7:

As a group, Hispanic candidates on the eligibility list resulting from Exam 7029 were ranked statistically significantly lower (i.e., had statistically significantly higher list numbers) than white candidates on the eligibility list. For purposes of this Request, “statistically significant” means that there is a 5% or lower likelihood that so large a disparity would occur by chance.

OBJECTION AND RESPONSE TO ADMISSION NO. 7:

Admits that a statistically significant difference exists between the ranked positions of Hispanic candidates, as a group, and the ranked positions of White candidates, as a group, on the eligibility list relating to Exam 7029 but defendant does not admit that any such difference was caused by the Exam.

ADMISSION NO. 8:

The disparity between the ranks of Hispanic candidates on the eligibility list resulting from Exam 7029 and the ranks of white candidates on the eligibility list is equivalent to more than three (3) units of standard deviation.

OBJECTION AND RESPONSE TO ADMISSION NO. 8:

Admits that the difference between the ranks of Hispanic candidates on the eligibility list relating to Exam 7029 and the ranks of White candidates on the eligibility list is equivalent to more than three (3) units of standard error so long as there are no controls for sample size, however when controls for sample size, such as those used by plaintiff’s expert, are

used in analyzing the rankings of these groups then the difference is less than three units of standard error. As explained in the appendix to defendant's expert report when comparing two groups, the appropriate measure of difference in ranking would be designated by "units of standard error," which although being a subset of "standard deviation" is not synonymous with the term "standard deviation."

ADMISSION NO. 9:

On average, the list numbers of black candidates on the eligibility list resulting from Exam 7029 were statistically significantly higher than the list numbers of white candidates on the eligibility list. For purposes of this Request, "statistically significant" means that there is a 5% or lower likelihood that so large a disparity would occur by chance.

OBJECTION AND RESPONSE TO ADMISSION NO. 9:

Admits that the statistical mean of the list numbers of Black candidates on the eligibility list relating to Exam 7029 was statistically significantly higher than the statistical mean of the list numbers of White candidates on the eligibility list but defendant does not admit that any such difference was caused by the Exam.

ADMISSION NO. 10:

The disparity between the list numbers of black candidates on the eligibility list resulting from Exam 7029 and the list numbers of white candidates on the eligibility list is equivalent to more than three (3) units of standard deviation.

OBJECTION AND RESPONSE TO ADMISSION NO. 10:

Admits that the difference between the list numbers of Black candidates on the eligibility list relating to Exam 7029 and the list numbers of White candidates on the eligibility

list is equivalent to more than three (3) units of standard error so long as there are no controls for sample size, however when controls for sample size, such as those used by plaintiff's expert, are used in analyzing the rankings of these groups then the difference is less than three units of standard error. As explained in the appendix to defendant's expert report when comparing two groups, the appropriate measure of difference in ranking would be designated by "units of standard error," which although being a subset of "standard deviation" is not synonymous with the term "standard deviation."

ADMISSION NO. 11:

On average, the list numbers of Hispanic candidates on the eligibility list resulting from Exam 7029 were statistically significantly higher than the list numbers of white candidates on the eligibility list. For purposes of this Request, "statistically significant" means that there is a 5% or lower likelihood that so large a disparity would occur by chance.

OBJECTION AND RESPONSE TO ADMISSION NO. 11:

Admits that the statistical mean of the list numbers of Hispanic candidates on the eligibility list relating to Exam 7029 was statistically significantly higher than the statistical mean of the list numbers of White candidates on the eligibility list so long as there are no controls for sample size, however when controls for sample size are used, such as those used by plaintiff's expert, in analyzing the statistical mean of the list numbers of these groups, then the difference is not statistically significant, but if any result has statistical significance, defendants do not admit that such a difference was caused by the Exam.

ADMISSION NO. 12:

The disparity between the list numbers of Hispanic candidates on the eligibility list resulting from Exam 7029 and the list numbers of white candidates on the eligibility list is equivalent to more than three (3) units of standard deviation.

OBJECTION AND RESPONSE TO ADMISSION NO. 12:

Admits that the difference between the ranks of Hispanic candidates on the eligibility list relating to Exam 7029 and the ranks of White candidates on the eligibility list is equivalent to more than three (3) units of standard error so long as there are no controls for sample size, however when controls for sample size are used, such as those used by plaintiff's expert, in analyzing the rankings of these groups then the difference is less than three units of standard error. As explained in the appendix to defendant's expert report when comparing two groups, the appropriate measure of difference in ranking would be designated by "units of standard error," which although being a subset of "standard deviation" is not synonymous with the term "standard deviation."

ADMISSION NO. 13:

The ratio of the pass rate of black candidates to the pass rate of white candidates on Written Exam 7029 is less than 80%.

OBJECTION AND RESPONSE TO ADMISSION NO. 13:

Admits that the ratio of the pass rate of Black candidates to the pass rate of White candidates on Written Exam 7029 is less than .80.

OBJECTION AND RESPONSE TO ADMISSION NO. 15:

The request is objected to on grounds that the phrase "effectively passed" is ambiguous, or rather, opaque. The "effective pass rate" could be determined by a number of methods which could result in varying "effective pass rates" So long as the means by which the determination of the "effective pass rate" is unclear, defendants do not have the ability to analyze the assertion. .

ADMISSION NO. 16:

The ratio of the rate at which black candidates on the eligibility list for Exam 2043 passed what Dr. Siskin refers to in his expert report as "the ranking test" to the rate at which white candidates on the eligibility list for Exam 2043 passed the ranking test is less than 80%. For purposes of this Request, "ranking test" means ranked at or above the rank of the last person appointed from the Exam 2043 eligibility list as of the date of the data provided to the United States.

OBJECTION AND RESPONSE TO ADMISSION NO. 16:

Admits that if the "the ranking test" used by Dr. Siskin in his expert report is applied to those persons on the eligibility list relating to Exam 2043, then the proportion of Black candidates on that eligibility list who passed Dr. Siskin's "ranking test" is less than 80% of the proportion of White candidates on that eligibility list who passed.

ADMISSION NO. 17:

Assuming that the City were to hire candidates down to list number 5,862 on the Exam 2043 eligibility list, the ratio of the rate at which black candidates on the eligibility list for Exam 2043 would pass what Dr. Siskin refers to in his expert report as "the ranking test" to the

rate at which white candidates on the eligibility list would pass the ranking test would be less than 80%. For purposes of this Request, "ranking test" means ranked at or above the rank of the last person appointed from the Exam 2043 eligibility list as of the date of the data provided to the United States.

OBJECTION AND RESPONSE TO ADMISSION NO. 17:

Admits that were the City to hire candidates as far down on the eligibility list relating to Exam 2043 as the candidate with the list number 5,862, as of the date that the data provided to the plaintiff, then the proportion of Black candidates passing what Dr. Siskin refers to in his expert report as "the ranking test," would be less than 80% of the proportion of White candidates passing what Dr. Siskin refers to in his expert report as "the ranking test."

ADMISSION NO. 18:

To the City's knowledge, no content validation study or analysis has been conducted for Written Exam 7029.

OBJECTION AND RESPONSE TO ADMISSION NO. 18:

Denied

ADMISSION NO. 19:

To the City's knowledge, no construct validation study or analysis has been conducted for Written Exam 7029.

OBJECTION AND RESPONSE TO ADMISSION NO. 19:

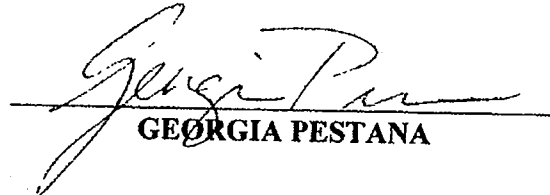
To the extent the request seeks an admission that construct validity methodology, as that term is used in the EEOC Guidelines, was absent from the validation study for Exam

VERIFICATION

GEORGIA PESTANA, declares under penalty of perjury pursuant to 28 U.S.C. § 1746, that the following is true and correct:

She has been duly designated as Assistant Corporation Counsel of The City of New York, and as such that she is an officer of The City of New York in the within action. That the foregoing Defendant's Amended Objections and Responses to Plaintiff's the United States' First Set Of Requests For Admission (Nos. 1-23) and Second Set Of Interrogatories (No. 33), are true to her knowledge except as to the matters therein stated to be alleged upon information and belief, and as to those matters she believes it to be true. Declarant further says that the reason why this verification is not made by The City of New York is that the City of New York is a corporation; that the grounds of her belief as to all matters is information obtained from the books and records of departments of the city government and from statements made to her by certain officers or agents of The City of New York.

Dated: New York, New York
March 5, 2008


GEORGIA PESTANA

OBJECTION AND RESPONSE TO ADMISSION NO. 27:

Admit.

ADMISSION NO. 28:

The City intended Written Exam 7029 to measure the following nine abilities (as defined on the page Bates numbered USA000397, which is attached to these Requests for Admission as Exhibit A): Written Comprehension, Written Expression, Memorization, Problem Sensitivity, Deductive Reasoning, Inductive Reasoning, Information Ordering, Spatial Orientation and Visualization.

OBJECTION AND RESPONSE TO ADMISSION NO. 28:

Admit.

ADMISSION NO. 29:

Each of the abilities Written Exam 7029 was intended to measure is cognitive ability.

OBJECTION AND RESPONSE TO ADMISSION NO. 29:

Admit.

ADMISSION NO. 30:

The City contends that Written Exam 7029 measures the following nine abilities (as defined on the page Bates numbered USA000397, which is attached to these Requests for Admission as Exhibit A): Written Comprehension, Written Expression, Memorization, Problem Sensitivity, Deductive Reasoning, Inductive Reasoning, Information Ordering, Spatial Orientation and Visualization.

OBJECTION AND RESPONSE TO ADMISSION NO. 30:

Admit.

ADMISSION NO. 31:

Written Exam 7029 is a paper and pencil examination consisting of 85 multiple choice questions.

OBJECTION AND RESPONSE TO ADMISSION NO. 31:

Admit.

ADMISSION NO. 32:

Although the City administered various versions of Written Exam 7029 (e.g., A.M., P.M., Sabbath Observer, etc.) at various times, each time the City administered Written Exam 7029, the examination items (i.e., questions) were the same, though the order of the items differed.

OBJECTION AND RESPONSE TO ADMISSION NO. 32:

Denied.

ADMISSION NO. 33:

Written Exam 2043 was developed by employees of the City, without the assistance or involvement of any outside (i.e., non-employee) expert(s) or consultants in the area of test development or validation.

OBJECTION AND RESPONSE TO ADMISSION NO. 33:

Denied.

ADMISSION NO. 34:

The City did not request the assistance or involvement of Catherine Cline, Ph.D., in the development of Written Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 34:

Admit.

ADMISSION NO. 35:

The City intended Written Exam 2043 to measure the following nine abilities (as defined on the page Bates numbered USA000397, which is attached to these Requests for Admission as Exhibit A): Written Comprehension, Written Expression, Memorization, Problem Sensitivity, Deductive Reasoning, Inductive Reasoning, Information Ordering, Spatial Orientation and Visualization.

OBJECTION AND RESPONSE TO ADMISSION NO. 35:

Admit.

ADMISSION NO. 36:

Each of the abilities Written Exam 2043 was intended to measure is cognitive ability.

OBJECTION AND RESPONSE TO ADMISSION NO. 36:

Admit.

ADMISSION NO. 37:

The City contends that Written Exam 2043 measures the following nine abilities (as defined on the page Bates numbered USA000397, which is attached to these Requests for Admission as Exhibit A): Written Comprehension, Written Expression, Memorization, Problem Sensitivity, Deductive Reasoning, Inductive Reasoning, Information Ordering, Spatial Orientation and Visualization.

OBJECTION AND RESPONSE TO ADMISSION NO. 37:

Admit.

ADMISSION NO. 38:

Written Exam 2043 is a paper and pencil examination consisting of 85 multiple choice questions.

OBJECTION AND RESPONSE TO ADMISSION NO. 38:

Admit.

ADMISSION NO. 39:

Although the City administered various versions of Written Exam 2043 (e.g., A.M., P.M., Sabbath Observer, etc.) at various times, each time the City administered Written Exam 2043, the examination items (i.e., questions) were the same, though the order of the items differed.

OBJECTION AND RESPONSE TO ADMISSION NO. 39:

Denied.

ADMISSION NO. 40:

The City used the same physical performance test ("PPT") for both Exam 7029 and Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 40:

Admit.

ADMISSION NO. 41:

For both Exam 7029 and Exam 2043, the PPT consisted of the same eight components or events.

OBJECTION AND RESPONSE TO ADMISSION NO. 41:

Admit.

ADMISSION NO. 42

For both Exam 7029 and Exam 2043, each of the components of the PPT was individually scored or timed.

OBJECTION AND RESPONSE TO ADMISSION NO. 42:

Denied.

OBJECTION AND RESPONSE TO ADMISSION NO. 38:

Admit.

ADMISSION NO. 39:

Although the City administered various versions of Written Exam 2043 (e.g., A.M., P.M., Sabbath Observer, etc.) at various times, each time the City administered Written Exam 2043, the examination items (i.e., questions) were the same, though the order of the items differed.

OBJECTION AND RESPONSE TO ADMISSION NO. 39:

Denied.

ADMISSION NO. 40:

The City used the same physical performance test ("PPT") for both Exam 7029 and Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 40:

Admit.

ADMISSION NO. 41:

For both Exam 7029 and Exam 2043, the PPT consisted of the same eight components or events.

OBJECTION AND RESPONSE TO ADMISSION NO. 41:

Admit.

ADMISSION NO. 42

For both Exam 7029 and Exam 2043, each of the components of the PPT was individually scored or timed.

OBJECTION AND RESPONSE TO ADMISSION NO. 42:

Denied.

ADMISSION NO. 43:

For each of the eight components of the PPT, the same passing standard was used for both Exam 7029 and Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 43:

Admit.

ADMISSION NO. 44:

For both Exam 7029 and Exam 2043, a candidate had to pass any six of the eight components comprising the PPT in order to pass the PPT.

OBJECTION AND RESPONSE TO ADMISSION NO. 44:

Admit.

ADMISSION NO. 45:

For both Exam 7029 and Exam 2043, the PPT score the City assigned to candidates who passed the PPT was the percentage of the eight PPT components the candidate passed.

OBJECTION AND RESPONSE TO ADMISSION NO. 45:

Admit.

ADMISSION NO. 46:

For both Exam 7029 and Exam 2043, there were only three possible passing scores on the PPT: 100, which was the score assigned candidates who passed all eight PPT components; 87.5, which was the score assigned candidates who passed seven of the eight PPT components; and 75, which was the score assigned candidates who passed six of the eight PPT components.

OBJECTION AND RESPONSE TO ADMISSION NO. 46:

Admit.

ADMISSION NO. 47:

For both Exam 7029 and Exam 2043, the City assigned a score of 62.5 to all candidates who failed the PPT.

OBJECTION AND RESPONSE TO ADMISSION NO. 47:

Admit.

ADMISSION NO. 48:

The application process for Exam 2043 was the same as the application process for Exam 7029, except that applications for Exam 2043 could also be obtained and submitted online.

OBJECTION AND RESPONSE TO ADMISSION NO. 48:

Admit.

ADMISSION NO. 49:

In order to be allowed to take Written Exam 7029, each applicant was required to return a completed application along with the required application fee (unless the fee was waived as set forth by City regulations) by mail to DCAS.

OBJECTION AND RESPONSE TO ADMISSION NO. 49:

Admit.

ADMISSION NO. 50:

In order to be allowed to take Written Exam 2043, each applicant was required to return a completed application along with the required application fee (unless the fee was waived as set forth by City regulations) to DCAS by mail or online.

OBJECTION AND RESPONSE TO ADMISSION NO. 50:

Admit.

ADMISSION NO. 51:

Each applicant for Exam 7029 who timely submitted a properly completed application and any required application fee was sent an admission card for Written Exam 7029.

OBJECTION AND RESPONSE TO ADMISSION NO. 51:

Admit.

ADMISSION NO. 52:

Each applicant for Exam 2043 who timely submitted a properly completed application and any required application fee was sent an admission card for Written Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 52:

Admit.

ADMISSION NO. 53:

While there were minimum qualifications for appointment for Exam 7029 and for Exam 2043, applicants were not screened for the applicable minimum qualifications prior to taking the examination.

OBJECTION AND RESPONSE TO ADMISSION NO. 53:

Admit.

ADMISSION NO. 54:

The City administered Written Exam 7029 to applicants for the first time on February 27, 1999.

OBJECTION AND RESPONSE TO ADMISSION NO. 54:

Denied.

ADMISSION NO. 55:

The City administered Written Exam 7029 for the last time in December 2002.

OBJECTION AND RESPONSE TO ADMISSION NO. 55:

Admit.

ADMISSION NO. 56:

The City administered Written Exam 2043 to applicants for the first time on December 14, 2002.

OBJECTION AND RESPONSE TO ADMISSION NO. 56:

Admit.

ADMISSION NO. 57:

The City continued to administer Written Exam 2043 at least as late as March 2007.

OBJECTION AND RESPONSE TO ADMISSION NO. 57:

Admit.

ADMISSION NO. 58:

For both Written Exam 7029 and Written Exam 2043, each candidate's score on the written examination was the percentage of the 85 questions on the examination that the candidate answered correctly.

OBJECTION AND RESPONSE TO ADMISSION NO. 58:

Admit.

ADMISSION NO. 59:

Before inviting a candidate who passed Written Exam 7029 to take the PPT, the City sent to the candidate a notice of results – i.e., a document that indicated whether the candidate had passed or failed the written examination and the candidate's score on the written examination.

OBJECTION AND RESPONSE TO ADMISSION NO. 59:

Admit.

ADMISSION NO. 60:

Before inviting a candidate who passed Written Exam 2043 to take the PPT, the City sent to the candidate a notice of results – i.e., a document that indicated whether the candidate had passed or failed the written examination and the candidate's score on the written examination.

OBJECTION AND RESPONSE TO ADMISSION NO. 60:

Admit.

ADMISSION NO. 61:

After a notice of results (i.e., a document that indicated whether the candidate passed or failed the written examination and the candidate's score on the written examination) was sent to a candidate who took Written Exam 7029, there was an appeals period during which the candidate could challenge the accuracy of his/her score on the written examination.

OBJECTION AND RESPONSE TO ADMISSION NO. 61:

Admit.

ADMISSION NO. 62:

After a notice of results (i.e., a document that indicated whether the candidate passed or failed the written examination and the candidate's score on the written examination) was sent to a candidate who took Written Exam 2043, there was an appeals period during which the candidate could challenge the accuracy of his/her score on the written examination.

OBJECTION AND RESPONSE TO ADMISSION NO. 62:

Admit.

ADMISSION NO. 63:

For both Exam 7029 and Exam 2043, a candidate's written examination score was corrected if the candidate timely submitted an appeal and it was determined that the original score was the result of a manifest error or a mistake.

OBJECTION AND RESPONSE TO ADMISSION NO. 63:

Admit.

ADMISSION NO. 64:

For both Exam 7029 and Exam 2043, if a candidate's written examination score was corrected as a result of an appeal, the candidate would be sent a new notice of results, reflecting the corrected written examination score.

OBJECTION AND RESPONSE TO ADMISSION NO. 64:

Admit.

ADMISSION NO. 65:

The City considered only those candidates who scored at or above 84.705 on Written Exam 7029 to have passed Written Exam 7029.

OBJECTION AND RESPONSE TO ADMISSION NO. 65:

Admit.

ADMISSION NO. 66:

The City considered only those candidates who scored above 70 on Written Exam 2043 to have passed Written Exam 2043.

OBJECTION AND RESPONSE TO ADMISSION NO. 66:

Denied.

ADMISSION NO. 67:

For both Exam 7029 and Exam 2043, the City sent to each candidate who passed the written examination a letter with instructions regarding the PPT, along with a description of the PPT.

OBJECTION AND RESPONSE TO ADMISSION NO. 67:

Denied.

ADMISSION NO. 68:

The City did not allow candidates who scored below 84.705 on Written Exam 7029 to take the PPT.

OBJECTION AND RESPONSE TO ADMISSION NO. 68:

Admit.

ADMISSION NO. 69:

The City did not allow candidates who scored below 70 on Written Exam 2043 to take the PPT.

OBJECTION AND RESPONSE TO ADMISSION NO. 69:

Admit.¹

ADMISSION NO. 70:

The Exam 7029 eligibility list was established in November 2000.

OBJECTION AND RESPONSE TO ADMISSION NO. 70:

Denied.

ADMISSION NO. 71:

The Exam 2043 eligibility list was established on May 5, 2004.

¹ Defendants note that DCAS typically extend their numbers to three decimal places. Thus, the City did not allow candidates who scored below 70.000 on Written Exam 2043 to take the PPT.

OBJECTION AND RESPONSE TO ADMISSION NO. 71:

Admit.

ADMISSION NO. 72:

The City placed all candidates who passed both Written Exam 7029 and the PPT on the Exam 7029 eligibility list.

OBJECTION AND RESPONSE TO ADMISSION NO. 72:

Denied.

ADMISSION NO. 73:

The City placed all candidates who passed both Written Exam 2043 and the PPT on the Exam 2043 eligibility list.

OBJECTION AND RESPONSE TO ADMISSION NO. 73:

Denied.

ADMISSION NO. 74:

On the Exam 7029 eligibility list, the City assigned each candidate a list number (or rank) based on the candidate's Adjusted Final Average, as calculated by the City, with the lowest list numbers (i.e., the highest ranks) being assigned to the candidates with the highest Adjusted Final Averages.

OBJECTION AND RESPONSE TO ADMISSION NO. 74:

Denied.

ADMISSION NO. 75:

On the Exam 2043 eligibility list, the City assigned each candidate a list number (or rank) based on the candidate's Adjusted Final Average, as calculated by the City, with the lowest list numbers (i.e., the highest ranks) being assigned to the candidates with the highest Adjusted Final Average.

ADMISSION NO. 84:

For Exam 7029, the City calculated each candidate's combined weighted standard score by multiplying the candidate's standardized written examination score by .5, multiplying the candidate's standardized PPT score by .5, rounding each of the products to five decimal places, adding the two products, and rounding to three decimal places.

OBJECTION AND RESPONSE TO ADMISSION NO. 84:

Admit.

ADMISSION NO. 85:

For Exam 7029, the City calculated each candidate's transformed score by multiplying the candidate's combined weighted standard score by 18.472906403940886699, rounding the product to three decimal places, adding 83.74384236453, and again rounding the result to three decimal places.

OBJECTION AND RESPONSE TO ADMISSION NO. 85:

Admit.

ADMISSION NO. 86:

For Exam 7029, each candidate's Adjusted Final Average was equal to the candidate's transformed score plus any applicable Residency, Legacy or Veterans' points.

OBJECTION AND RESPONSE TO ADMISSION NO. 86:

Admit.

ADMISSION NO. 87:

The City calculated the Adjusted Final Average of each candidate on the Exam 2043 eligibility list as set forth in the document Bates numbered USA004931-USA004932, which is attached to these Requests for Admission as Exhibit C.

OBJECTION AND RESPONSE TO ADMISSION NO. 95:

Admit.

ADMISSION NO. 96:

For Exam 2043, each candidate's Adjusted Final Average was equal to the candidate's transformed score plus any applicable Residency, Legacy or Veterans' points.

OBJECTION AND RESPONSE TO ADMISSION NO. 96:

Admit.

ADMISSION NO. 97:

In order to be appointed from the Exam 7029 or Exam 2043 eligibility list, a candidate had to appear on a certification list then in force and effect and meet all requirements for appointment set forth in the relevant notice of examination.

OBJECTION AND RESPONSE TO ADMISSION NO. 97:

Denied.

ADMISSION NO. 98:

The Candidate Investigation Division ("CID") of the FDNY conducted the processing of candidates from the Exam 7029 and Exam 2043 eligibility lists to determine whether the candidates meet the requirements for appointment.

OBJECTION AND RESPONSE TO ADMISSION NO. 98:

Admit.

ADMISSION NO. 99:

CID did not necessarily wait until a candidate appeared on a certification list to begin processing the candidate (i.e., to begin the process of determining whether a candidate on the eligibility list was qualified for appointment).

ADMISSION NO. 103:

For Exam 7029 and Exam 2043, CID did not assign candidates to investigators for background investigation until after the candidate's intake interview.

OBJECTION AND RESPONSE TO ADMISSION NO. 103:

Admit.

ADMISSION NO. 104:

For Exam 7029 and Exam 2043, CID did not assign candidates to investigators for background investigation until it believed the candidate's list number would likely be reached for possible appointment to the next academy class.

OBJECTION AND RESPONSE TO ADMISSION NO. 104:

Admit.

ADMISSION NO. 105:

The number of candidates that were certified from the Exam 7029 or Exam 2043 eligibility list at a given time was based on an assessment by the DCAS Certification Unit and the FDNY of how many candidates would be needed on the certification list in order to fill an upcoming academy class, given the number of candidates who appeared to be "not qualified" at that time based on CID's ongoing investigations.

OBJECTION AND RESPONSE TO ADMISSION NO. 105:

Admit.

ADMISSION NO. 106:

For Exams 7029 and 2043, a candidate would have been considered "not qualified" on a given certification list if CID had been unable to contact the candidate to begin evaluating the candidate's qualifications, if the candidate had not yet provided documentation of his or her qualifications, if the candidate had not yet completed requirements for appointment

which could later be completed (such as education or, when required for appointment, CFR-D certification), or if the candidate had not otherwise completed the steps in the selection process.

OBJECTION AND RESPONSE TO ADMISSION NO. 106:

Denied.

ADMISSION NO. 107:

A candidate who was deemed not qualified on one certification list could be considered qualified and appointed from a later certification list.

OBJECTION AND RESPONSE TO ADMISSION NO. 107:

Admit.

ADMISSION NO. 108:

Candidates on the Exam 7029 and Exam 2043 eligibility lists were allowed to withdraw temporarily from consideration.

OBJECTION AND RESPONSE TO ADMISSION NO. 108:

Admit.

ADMISSION NO. 109:

A candidate on the Exam 7029 or Exam 2043 eligibility list who temporarily withdrew from consideration could be certified for the first time later than lower ranked candidates (i.e., candidates with higher list numbers).

OBJECTION AND RESPONSE TO ADMISSION NO. 109:

Admit.

ADMISSION NO. 110:

Each certification list began with the highest ranked candidate (i.e., the candidate with lowest list number) on the eligibility list from which the certification list was drawn.

OBJECTION AND RESPONSE TO ADMISSION NO. 110:

Admit.

ADMISSION NO. 111:

Each certification list consisted of candidates drawn in rank (i.e., list number) order from the eligibility list, beginning with the highest-ranking candidate who had not withdrawn from consideration and who remained eligible for certification.

OBJECTION AND RESPONSE TO ADMISSION NO. 111:

Admit.

ADMISSION NO. 112:

The City appointed candidates on certification lists drawn from the Exam 7029 or Exam 2043 eligibility list in descending rank order (i.e., ascending list number order) from among those candidates on the certification list who had completed the investigation process, had been determined to be qualified at the time a new academy class was appointed, and accepted an offer of employment.

OBJECTION AND RESPONSE TO ADMISSION NO. 112:

Admit.

ADMISSION NO. 113:

If the list number of a candidate on the Exam 7029 or 2043 eligibility list had not been reached by the date on which the last appointment to a given academy class was made (i.e., if all candidates appointed to the class from the eligibility list had lower list numbers), the candidate was not appointed at that time even if the candidate had completed all steps in the selection process and been found qualified.

OBJECTION AND RESPONSE TO ADMISSION NO. 113:

Admit.

OBJECTION AND RESPONSE TO ADMISSION NO. 121:

Denied.

ADMISSION NO. 122:

All entry-level firefighters appointed from an eligibility list resulting from an open competitive selection process since May 2008 have been appointed from the Exam 6019 eligibility list.

OBJECTION AND RESPONSE TO ADMISSION NO. 122:

Denied.

ADMISSION NO. 123:

The Exam 6019 eligibility list was established on June 11, 2008.

OBJECTION AND RESPONSE TO ADMISSION NO. 123:

Admit.

ADMISSION NO. 124:

The City first administered Written Exam 6019 on January 20, 2007.

OBJECTION AND RESPONSE TO ADMISSION NO. 124:

Admit.

ADMISSION NO. 125:

Although the City administered various versions of Written Exam 6019 (e.g., A.M., P.M., Sabbath Observer, etc.) at various times, each time the City administered Written Exam 6019, the examination items (i.e., questions) were the same, though the order of the items differed.

OBJECTION AND RESPONSE TO ADMISSION NO. 125:

Denied.

ADMISSION NO. 126:

Written Exam 6019 is a paper and pencil examination, consisting of 150 multiple choice questions.

OBJECTION AND RESPONSE TO ADMISSION NO. 126:

Admit.

ADMISSION NO. 127:

Written Exam 6019 is "objectively" scored, as that term is used in the expert report of the City's experts, Drs. Philip Bobko and F. Mark Schemmer.

OBJECTION AND RESPONSE TO ADMISSION NO. 127:

Admit.

ADMISSION NO. 128:

The City intended Written Exam 6019 to measure the following 20 abilities and personal attributes (as defined in the Abilities List Bates numbered DCAS-0008687 to DCAS-0008695, which is attached to these Requests for Admission as Exhibit D): Written Comprehension, Memorization, Problem Sensitivity, Number Facility, Deductive Reasoning, Inductive Reasoning, Information Ordering, Speed of Closure, Flexibility of Closure, Spatial Orientation, Visualization, Perceptual Speed, Time Sharing, Adaptability, Tenacity, Integrity, Work Standards, Resilience, Coordination, and Establishing and Maintaining Interpersonal Relationships.

OBJECTION AND RESPONSE TO ADMISSION NO. 128:

Denied.

ADMISSION NO. 129:

Some of the abilities and personal attributes the City intended Written Exam 6019 to measure are cognitive.

VERIFICATION

GEORGIA PESTANA, declares under penalty of perjury pursuant to 28 U.S.C. § 1746, that the following is true and correct:

She has been duly designated as Assistant Corporation Counsel of The City of New York, and as such that she is an officer of The City of New York in the within action. That the foregoing Defendant's Objections and Responses to Plaintiff's the United States' Second Set Of Requests For Admission (Nos. 24-132) and Second Set Of Interrogatories (No. 36), are true to her knowledge except as to the matters therein stated to be alleged upon information and belief, and as to those matters she believes it to be true. Declarant further says that the reason why this verification is not made by The City of New York is that the City of New York is a corporation; that the grounds of her belief as to all matters is information obtained from the books and records of departments of the city government and from statements made to her by certain officers or agents of The City of New York.

Dated: New York, New York
September 8, 2008

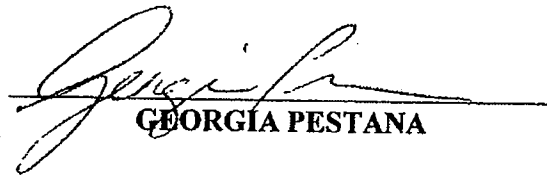

GEORGIA PESTANA

EXHIBIT N

objections and construing this interrogatory as relating only to Exam Nos. 7029 and 2043, defendant refers plaintiff to USA 4811-4918, USA 4623-4680.

INTERROGATORY NO. 15:

For each written examination administered since January 1, 1999 by the City as part of an open competitive examination process for appointment to the rank of entry-level firefighter in the FDNY, identify and state the race/national origin, position and telephone number of each person involved in the development, validation or selection of the written examination; state which written examination(s) the individual was involved in developing, validating or selecting; and describe the nature of each such person's involvement. For purposes of this interrogatory, "selection" of the examination includes any decision regarding whether the examination would be developed in-house by DCAS.

OBJECTION AND RESPONSE TO INTERROGATORY NO. 15:

Defendant repeats its objection and response to interrogatory 14 and also refers plaintiff to USA 4680-4718.

INTERROGATORY NO. 16:

For each written examination administered since January 1, 1999 by the City as part of an open competitive examination process for appointment to the rank of entry-level firefighter in the FDNY, if the City used the written examination as a pass/fail screening device (regardless whether the City also used the written examination in any other manner): identify and state the race/national origin, position and telephone number of each person involved in determining (or in any study or analysis done to determine) the cutoff score the City would use; describe the nature of each such person's involvement; state when the cutoff score was determined (if exact date is not known, state approximate date or indicate the point in the open competitive examination process (e.g., after scoring of written examinations)); state when and

how applicants were notified whether they passed or failed; and state the reason(s) the City decided to use the particular cutoff score the City used.

OBJECTION AND RESPONSE TO INTERROGATORY NO. 16:

Defendant objects to this interrogatory on the ground that it is duplicative and cumulative, states that it is searching for information with respect to the cut-off score established for the written portion of Exam 7029 and that the cut-off score for Exam 2043 was set pursuant to DCAS Personnel Rule and Regulation 4.4.9 (see USA 1188-1196).

INTERROGATORY NO. 17:

For each written examination administered since January 1, 1999 by the City as part of an open competitive examination process for appointment to the rank of entry-level firefighter in the FDNY: identify and state the race/national origin, position, and telephone number of each person involved in determining (or in any study or analysis done to determine) how the City would rank-order applicants on the eligibility list resulting from the examination (e.g., whether the City would rank-order applicants on the basis of written examination score or combined written examination and PPT scores (plus bonus points) and/or any formula(s) use to standardize or combine written examination and PPT scores; describe the nature of each such person's involvement; and state the reason(s) the City decided to rank-order applicants in the manner in which it did (including, but not limited to, the reason(s) the City decided to weight written examination and PPT scores in the manner in which it did).

OBJECTION AND RESPONSE TO INTERROGATORY NO. 17:

Defendant objects to this interrogatory to the extent that it is cumulative and duplicative of document requests and on the ground that the information is obtainable from another source that is more convenient and less burdensome, specifically, the deposition of Thomas Patitucci and William Klimowitz.

INTERROGATORY NO. 30:

Separately for Written Exam 7029 and Written Exam 2043, state whether the City contends that the cutoff score used on the examination measures the minimum level of the tested skills, abilities or other characteristics necessary for successful performance of the job of entry-level firefighter in the FDNY; and, if so, state all facts and identify all documents relating to the City's contention.

SUPPLEMENTAL OBJECTION AND RESPONSE TO INTERROGATORY NO. 30:

Defendant objects to this interrogatory as vague and ambiguous. Notwithstanding this objections, the City makes no such contention.

INTERROGATORY NO. 31:

Separately for Written Exam 7029 and Written Exam 2043, state whether the City contends that rank-order processing on the basis of combined written test score and PPT score (combined in the manner in which the City combined them) is/was job related and consistent with business necessity within the meaning of 42 U.S.C. §§2000ee-2(k)(1)(A)(i); and, if so, state all facts and identify all documents relating to the City's contention.

SUPPLEMENTAL OBJECTION AND RESPONSE TO INTERROGATORY NO. 31:

Defendant objects to this interrogatory as vague and ambiguous. Notwithstanding this objection, the City responds in the affirmative.

Facts Concerning the Job relatedness of Plaintiff of Exams 7029 and 2043 are reflected in the documents bearing the Bate stamp numbers: USA004811 to USA004918; USA000329 to USA000404 which are the Test Development Report for Exam 7029; USA004357 to USA004918 which are documents relating to the Linking Panel for Exam 7029; USA004937, USA004938 to USA006229 which are Item Records for Exam 7029; USA006432

INTERROGATORIES

INTERROGATORY NO. 36:

For each Request for Admission contained in Plaintiff United States' Second Set of Requests for Admission (Nos. 24 - 132) which the City denies in whole or in part, state all facts and identify all documents on which the City's denial is based or otherwise relating to the City's denial and identify all persons with knowledge of the facts stated or the location of the documents identified.

OBJECTION AND RESPONSE TO INTERROGATORY NO. 36:

Defendant objects to this request to the extent it seeks disclosure of attorney work-product and attorney-client communications. Notwithstanding the forgoing objections, Defendant states that:

Request for Admission Number 24 is denied because Defendants do not concede jurisdiction of the Court over the claims of the Plaintiff-Intervenors.

Request for Admission Number 26 is denied because the City, in developing Written Exam 7029, relied, in part, on the work Dr. Landy performed in the development of Written Exam 0084.

Request for Admission Number 32 is denied because although the City administered various versions of Written Exam 7029 (e.g., A.M., P.M., Sabbath Observer, etc.) at various times, and at each administration of Written Exam 7029 the examination items (i.e., questions) were the same, sometimes, the order of the items, and/or individual item options (i.e., individual answer choices) differed.

Request for Admission Number 33 is denied because the City, in developing Written Exam 7029, relied, in part, on the work Dr. Landy performed in the development of

Written Exam 0084. Similarly, the City relied on Written Exam 7029 in developing Written Exam 2043.

Request for Admission Number 39 is denied because although the City administered various versions of Written Exam 2043 (e.g., A.M., P.M., Sabbath Observer, etc.) at various times, and at each administration of Written Exam 2043 the examination items (i.e., questions) were the same, sometimes, the order of the items, and/or individual item options (i.e., individual answer choices) differed.

Request for Admission Number 42 is denied because for both Exam 7029 and 2043 each of the components of the PPT were individually timed, but not individually scored.

Request for Admission Number 54 is denied because the City administered Written Exam 7029 to applicants for the first time on February 26, 1999.

Request for Admission Number 66 is denied because the City considered only those candidates who scored at or above 70.000 on Written Exam 2043 to have passed Written Exam 2043.

Request for Admission Number 67 is denied because for both Exam 7029 and 2043, the City sent to each candidate who passed the written examination a packet with instructions regarding the PPT.

Request for Admission Number 70 is denied because the Exam 7029 eligibility list was established on November 15, 2000.

Request for Admission Number 72 is denied because there may be instances where a candidate passed the PPT and Written Exam 7029, but had not corrected defects in their paperwork at the time of the establishment of the list. Thus, such a candidate who passed the PPT and Written Exam 7029 could not be placed on the list.

Request for Admission Number 73 is denied because there may be instances where a candidate passed the PPT and Written Exam 2043, but had not corrected defects in their paperwork at the time of the establishment of the list. Thus, such a candidate who passed the PPT and Written Exam 2043 could not be placed on the list.

Request for Admission Number 74 is denied because on the Exam 7029 eligibility list, the City could not assign list numbers (or ranks) to candidates who passed the exam based on their Adjusted Final Average, as calculated by the City, unless and until the candidate had rectified all potential deficiencies in their paperwork.

Request for Admission Number 75 is denied because on the Exam 2043 eligibility list, the City could not assign list numbers (or ranks) to candidates who passed the exam based on their Adjusted Final Average, as calculated by the City, unless and until the candidate had rectified all potential deficiencies in their paperwork.

Request for Admission Number 76 is denied because candidates who had the same Adjusted Final Average were ranked (i.e., assigned list numbers) on the eligibility list based on the last five digits of the candidates' Social Security Numbers.

Request for Admission Number 97 is denied because in order to be appointed from the Exam 7029 or Exam 2043 eligibility list, a candidate had to appear on the certification list then in effect and meet all requirements for appointment set forth in the relevant Notice of Examination and pass a medical and psychological evaluation.

Request for Admission Number 99 is denied because CID was not obligated to, but could, begin processing the candidate (i.e., to begin the process of determining whether a candidate on the eligibility list was qualified for appointment) prior to reaching their name on the certification list.

Request for Admission Number 100 is denied because in order to determine whether candidates from the Exam 7029 and 2043 eligibility lists met the requirements for appointment, CID investigated candidates' character and background, which investigation included examining criminal record, military history, references, educational background, CFR-D certification (when required from appointment), English-speaking ability, citizenship, age, proof of identification and driving record.

Request for Admission Number 106 is denied because if CID had been unable to contact a candidate, the candidate would be listed on the certification list as Failed To Report ("FRI"). If the candidate had not completed all of the requirements for appointment, the candidate would be listed as Not Qualified for Appointment ("NQA"). Finally, if the candidate had not completed all of the steps in the process, they would be listed as FRI and in the remarks would be Failed To Cooperate ("FTC").

Request for Admission Number 117 is objected to on the grounds that as phrased the requested is confusing and difficult to comprehend. Notwithstanding and subject to this objection, and the objections listed in Defendants' General Statement, Defendants admit that a candidate was deemed "not qualified" as opposed to "considered not selected," if CID was unable to contact the candidate to begin evaluating the candidate's qualifications, or if the candidate had not yet provided documentation of his or her qualifications, or if the candidate had not yet completed requirements for appointment which could alter be completed (such as education or CFR-D certification), or if the candidate had not otherwise completed the steps in the selection process.

Request for Admission Number 119 is denied because a new certification list could not be issued until the previous certification list drawn from the same eligibility list was returned and processed.

Request for Admission Number 120 is denied because the City appointed entry level firefighters from the Exam 7029 eligibility list between February 2001 until September 2004. The list terminated on November 15, 2004. There were no appointments made from the list after November 15, 2004. However, appointments continue to be made from the special military list associated with Exam 7029.

Request for Admission Number 121 is denied because the City appointed entry level firefighters from the Exam 2043 eligibility list from May 2004 until January 2008. The list terminated on May 5, 2008. There were no appointments made from the list after May 5, 2008. However, appointments continue to be made from the special military list associated with Exam 2043.

Request for Admission Number 122 is denied because the Exam 6019 eligibility list was not established until June 11, 2008. This list is expected to be in existence for four years. The first certification upon which the names of candidates from Exam 6019 appear was issued on June 25, 2008; that certification, however, included names from special military lists associated with Exams 7029 and 2043.

Request for Admission Number 125 is denied because although the City administered various versions of Written Exam 6019 (e.g., A.M., P.M., Sabbath Observer, etc.) at various times, and at each administration of Written Exam 6019 the examination items (i.e., questions) were the same, sometimes, the order of the items, and/or individual item options (i.e., individual answer choices) differed.

Request for Admission Number 128 is objected to based on plaintiff's use of the term "personal attributes," which does not appear in the Exam 6019 Job Analysis Report.

Request for Admission Number 129 is objected to based on plaintiff's use of the term "personal attributes," which does not appear in the Exam 6019 Job Analysis Report.

Request for Admission Number 130 is objected to based on plaintiff's use of the term "personal attributes," which does not appear in the Exam 6019 Job Analysis Report.

Request for Admission Number 131 is objected to based on plaintiff's use of the term "personal attributes," which does not appear in the Exam 6019 Job Analysis Report.

Dated: New York, New York
September 8, 2008

MICHAEL A. CARDOZO
Corporation Counsel of the City of New York
Attorney for Defendant
100 Church Street, Room 2-100
New York, New York 10007
(212) 788-1247
wfraenke@law.nyc.gov

By: William S.J. Fraenkel
WILLIAM S.J. FRAENKEL
Assistant Corporation Counsel

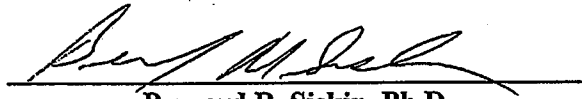
EXHIBIT O

REPORT

in the matter of

United States of America v. City of New York

by


Bernard R. Siskin, Ph.D.
Director and Head of Labor Practice Group

**LECG
Philadelphia PA**

November 2007

-3-

resulted in a statistically significant disparate impact upon African-American candidates for the position of entry-level firefighter in the FDNY. The disparity between the pass rate of African-American candidates and the pass rate of white candidates is equivalent to 33.9 units of standard deviation.

- The City's pass/fail use of Written Exam 7029 also has resulted in a statistically significant disparate impact upon Hispanic candidates for the position of entry-level firefighter in the FDNY. The disparity between the pass rates of Hispanic and white candidates is equivalent to 17.41 units of standard deviation.
- In practical terms, the effect of the disparate impact of the City's pass/fail use of Written Exam 7029 upon African Americans was to eliminate 519 African-American candidates from any possibility of appointment as entry-level firefighters in the FDNY. In other words, had the City's pass/fail use of Written Exam 7029 not had a disparate impact, 519 (74.7%) of the African Americans who failed Written Exam 7029 would have passed it, and an estimated 114 more African Americans would have been appointed as firefighters. The pass rate of African-American candidates on Written Exam 7029 was 67.0% of the pass rate of white candidates.
- In practical terms, the effect of the disparate impact of the City's pass/fail use of Written Exam 7029 upon Hispanics was to eliminate 282 Hispanic candidates from any possibility of appointment as entry-level firefighters in the FDNY. Had the City's pass/fail use of Written Exam 7029 not had a disparate impact, 282 (56.9%) of the Hispanics who failed Written Exam 7029 would have passed it, and an estimated 62 more Hispanics would have been appointed as firefighters. The pass

-4-

rate of Hispanic candidates on Written Exam 7029 was 85.3% of the pass rate of white candidates.

B. The City's Rank-Order Processing/Selection of Candidates from the Exam 7029 Eligibility List Resulted in a Disparate Impact upon Both African Americans and Hispanics

- African Americans who passed both Written Exam 7029 and the PPT were ranked statistically significantly lower on the eligibility list for Exam 7029 than whites. The disparity in ranks between African Americans and whites is equivalent to 6.48 units of standard deviation.
- Hispanics who passed both Written Exam 7029 and the PPT were ranked statistically significantly lower on the eligibility list for Exam 7029 than whites. The disparity in ranks between Hispanics and whites is equivalent to 4.57 units of standard deviation.
- In practical terms, the effect of the disparate impact upon African Americans of the City's rank-order processing/selection from the Exam 7029 eligibility list was that, on average, African Americans were delayed in being reached on the eligibility list and, hence, were delayed in being appointed. The delay resulted in a loss of wages and seniority for 68 African-American firefighters appointed from Exam 7029. These 68 African-American firefighters lost a total of approximately 20 years of employment. On average, each of the 68 lost about 3 1/2 months of wages and seniority due to the disparate impact of the rank-order selection process.
- In practical terms, the effect of the disparate impact upon Hispanics of the City's rank-order processing/selection from the Exam 7029 eligibility list was that, on

-5-

average, Hispanics were delayed in being reached on the eligibility list and, hence, were delayed in being appointed. The delay resulted in a loss of wages and seniority for 86 Hispanic firefighters appointed from Exam 7029. These 86 Hispanic firefighters lost a total of approximately 23 years and 1 month of employment. On average, each of the 86 lost more than 3 months of wages and seniority due to the disparate impact of the rank-order selection process.

C. The City's Pass/Fail Use of Written Exam 2043 with a Cutoff Score of 70 Has Resulted in a Disparate Impact upon Both African-American and Hispanic Candidates

- The City's pass/fail use of Written Exam 2043 with a cutoff score of 70 has resulted in a statistically significant disparate impact upon African-American candidates for the position of entry-level firefighter in the FDNY. The disparity between the pass rates of African-American and white candidates is equivalent to 21.84 units of standard deviation.
- The City's pass/fail use of Written Exam 2043 with a cutoff score of 70 has resulted in a statistically significant disparate impact upon Hispanic candidates for the position of entry-level firefighter in the FDNY. The disparity between the pass rates of Hispanic and white candidates is equivalent to 10.46 units of standard deviation.
- In practical terms, the effect of the disparate impact of the City's pass/fail use of Written Exam 2043 upon African Americans was to eliminate 165 African-American candidates from any possibility of appointment. Had the City's pass/fail use of Written Exam 2043 not had a disparate impact, 165 (81.3%) of the African Americans who failed Written Exam 2043 would have passed it, and an estimated 30 of them

-10-

list has resulted in a disparate impact upon both African-American and Hispanic candidates, resulting in delayed appointments for 154 of the African-American and Hispanic firefighters who were appointed from Exam 7029, for a total loss of over 43 years of wages and seniority by these African-American and Hispanic firefighters.

- In summary, the City's pass/fail use of Written Exam 2043 also has resulted in a disparate impact upon both African-American and Hispanic candidates. As a result of this disparate impact, 259 fewer African-American and Hispanic candidates passed Written Exam 2043, and I estimate that, to date, the City has appointed 47 fewer African-American and Hispanic firefighters than it would have absent the disparate impact of its pass/fail use of Written Exam 2043. In addition, the City's rank-order processing/selection from the Exam 2043 eligibility list has resulted in a disparate impact upon both African-American and Hispanic candidates. As a result of this disparate impact, to date: (i) 158 fewer African-American and Hispanic candidates have been reached on the eligibility list for possible appointment, and I estimate that the City has appointed 70 fewer African-American and Hispanic firefighters than it would have absent the disparate impact of its rank-order process; and (ii) the appointments of 95 of the African-American and Hispanic firefighters who have been appointed from Exam 2043 have been delayed, for a total loss of over 26 years of wages and seniority by African-American and Hispanic firefighters.

III. DATA

My analyses in this report rely upon two databases for each Exam (7029 and 2043) supplied by the City. One database, the applicant data file, contains the name, race/national origin, Social

-12-

conclusions regarding the disparate impact of the practices challenged by the United States. See Appendix C for a description of the data inconsistencies and my handling of them for purposes of my analyses.

IV. BACKGROUND AND METHODOLOGY

A. Overview of Examination Process

Since February 2001, the City has appointed entry-level firefighters from two open competitive examination processes, Exam 7029 and Exam 2043. Exams 7029 and 2043 each involved the administration of a written examination as well as a PPT. The written examination used as part of Exam 7029 was administered in February 1999.² The City used the Exam 7029 eligibility list to appoint firefighters from February 2001 until December 2004. The written examination used as part of Exam 2043 was administered in December 2002.³ The City has used the Exam 2043 eligibility list since May 2004 (*i.e.*, there was some overlap in the City's use of the Exam 7029 and 2043 eligibility lists) and has stated that it will continue to use the Exam 2043 eligibility list until May 2008, when the list will expire.

The City used both Written Exam 7029 and Written Exam 2043 on a "pass/fail basis." The

² It is my understanding that the main administration of Written Exam 7029 took place on February 27, 1999, and most candidates took the written examination on that date, but the City also administered Written Exam 7029 on other dates. The earliest written examination date in the applicant data produced by the City for Exam 7029 is February 26, 1999. The latest written examination date in the applicant data produced by the City for Exam 7029 is December 14, 2002.

³ It is my understanding that the first administration of Written Exam 2043 took place on December 14, 2002, and most candidates took the written examination on that date, but the City continues to administer Written Exam 2043. The latest written examination date in the Exam 2043 applicant data produced by the City last September is March 23, 2007.

-15-

the passing rate for white candidates as the benchmark because the *Uniform Guidelines on Employee Selection Procedures* suggest that the success rate of the class of interest (African Americans or Hispanics) should be compared to that of the best performing or majority group (whites). If there is a statistically significant difference between the pass rates of African-American and white candidates, or Hispanic and white candidates, one can conclude that this disparity is associated with race or national origin, respectively, rather than chance. In other words, if African-American or Hispanic candidates pass the written examination at a statistically significantly lower rate than whites, then the adverse impact on African Americans and Hispanics is not the result of chance but, instead, is attributable to the fact that African-American and Hispanic candidates really have a lower likelihood of passing the written examination than do white candidates.

Statisticians, testing experts, and other social scientists normally consider a disparity to be statistically significant if there is a 5% or lower likelihood (*i.e.*, probability) that so large a disparity would occur by chance. A disparity can also be described in terms of units of standard deviation. Units of standard deviation correspond to the likelihood that an observed disparity at least as large as the one obtained would occur by chance. For example, the 5% standard equates to slightly less than two (1.965) units of standard deviation. A disparity of three units of standard deviation equates to a 0.3% (*i.e.*, 3-in-1,000) likelihood of at least as large a disparity occurring by chance.

In some instances, a statistically significant disparity may not have a real-world effect large enough to be considered of practical significance. That is ultimately a judgment that must be made by the trier of fact; it cannot be determined by a statistical test. However, descriptive statistics can assist the decision maker by providing information that quantifies the practical effect of the disparity. One such statistic is the shortfall in African-American or Hispanic candidates who pass the

-21-

Americans and Hispanics the same as the overall distribution (by moving African Americans and Hispanics up on the list), and (ii) calculating the lost period of employment (and, hence, wages and seniority) of African Americans and Hispanics that resulted from the disparate impact of the ranking process.¹³

V. ANALYSIS AND FINDINGS

A. Pass/Fail Use of Written Exam 7029

1. The City's Pass/Fail Use of Written Exam 7029 with a Cutoff Score of 84.705 Resulted in a Disparate Impact upon African-American Candidates

As stated previously, the City used a passing score of 84.705 for Written Exam 7029. Only candidates who passed Written Exam 7029 were allowed to continue on in the selection process and take the PPT. Table 1 summarizes the disparate impact of the City's pass/fail use of Written Exam 7029 upon African Americans.

Table 1 shows that the City's pass/fail use of Written Exam 7029 had a highly statistically significant disparate impact upon African-American candidates. The disparity is equivalent to 33.9 units of standard deviation. The likelihood of such a disparity occurring by chance is infinitesimally small, less than 1-in-4.5 million billion.

I understand that the City has suggested that this highly statistically significant finding is due to the large sample size involved in the analysis of FDNY firefighter candidate data. The City is correct that a given disparity in pass rates may be statistically significant with a large sample and not with a smaller sample. In other words, the larger the sample size, the smaller the actual disparity need

¹³ These computations use the actual appointment dates. Any deviations in appointment dates from perfect rank order (see footnote 6 for why this might occur) are assumed to be independent of race and national origin. If the analysis were to assume that all appointments were made in perfect rank order, there would be little difference in the results.

-22-

be to reach statistical significance. Thus, a finding of statistical significance does in part reflect the large sample size. However, the finding of statistical significance does not depend only on sample size; it also is a factor of the size of the disparity. Here, if the Exam 7029 sample size were reduced by 90% (so that there were only 1,292 white candidates and 175 African-American candidates) and the same disparity between the pass rates of African-American candidates and white candidates were to occur, the results still would be highly statistically significant. With the reduced sample size, the disparity would be equivalent to 10.53 units of standard deviation. The likelihood of such a disparity occurring by chance is infinitesimally small, less than 1-in-4.5 million billion. Thus, the City's suggestion that a significant disparity would not have been found if the sample had not been so large is not supported by the data.

In terms of practical effect, Table 1 shows that the disparate impact of the City's pass/fail use of Written Exam 7029 resulted in a shortfall of 519 African-American written examination passers. That is, if African-American candidates had passed Written Exam 7029 at the same rate as white candidates, an additional 519 African Americans, or nearly three-quarters (74.5%) of the African Americans who failed the written examination, would have passed it. Using the methodology described in Section IV. B. of my report, an estimated 114 of them would have been appointed.¹⁴ Thus, if the test had not resulted in a disparate impact upon African Americans, the number of African-American appointees would have increased by over 100%, from 104 to 218. The ratio of the African-American pass rate to the white pass rate is 67.0%.

¹⁴ Since the eligibility list was exhausted, the only reason a candidate on the eligibility list would not have been appointed is because the candidate was disqualified or unavailable, not because of a lack of available positions. The estimation methodology accounts for candidates on the list being disqualified or unavailable.

-23-

2. The City's Pass/Fail Use of Written Exam 7029 with a Cutoff Score of 84.705 Resulted in a Disparate Impact upon Hispanic Candidates

Table 2 summarizes the disparate impact of the City's pass/fail use of Written Exam 7029 on Hispanics. Table 2 shows that the City's use of Written Exam 7029 with a cutoff score of 84.705 had a highly statistically significant disparate impact upon Hispanic candidates. The disparity between the Hispanic and white pass rates is equivalent to 17.41 units of standard deviation. The likelihood of such a disparity occurring by chance is infinitesimally small, less than 1-in-4.5 million billion.

Again, I understand that the City suggests that this highly statistically significant finding is due to the large sample size. As noted above, although the City is correct that the larger the sample size, the smaller the actual disparity must be to reach statistical significance, the City fails to consider that my finding of statistical significance also depends on the size of the disparity. If the sample size were 90% smaller (so that there were only 1,292 white candidates and 213 Hispanic candidates) and the same disparity were to occur, the results still would be highly statistically significant. With the reduced sample size, the disparity would be equivalent to 5.48 units of standard deviation (an event which would occur by chance less than 1-in-24 million times). Thus, the finding of statistical significance cannot reasonably be attributed to the large sample size.

In terms of practical effect, Table 2 shows that, as a result of the City's pass/fail use of Written Exam 7029, there was a shortfall of 282 Hispanic written examination passers. In other words, more than one-half (56.9%) of the 496 Hispanic candidates who failed Written Exam 7029 would have passed it if there had not been a disparate impact upon Hispanics. Using the methodology described in Section IV. B. of my report, an estimated 62 of them would have been appointed. Thus, if the test had not resulted in a disparate impact upon Hispanics, the number of Hispanic appointees would have

-24-

increased by about 23%, from 273 to 335. The ratio of the Hispanic pass rate to the white pass rate is 85.3%.

B. Rank-Order Processing/Selection of Candidates from the Exam 7029 Eligibility List Based upon Combined Written Examination and PPT Scores

1. The City's Rank-Order Processing/Selection of Candidates from the Exam 7029 Eligibility List Resulted in a Disparate Impact upon African Americans

As explained previously, candidates who passed both Written Exam 7029 and the PPT were placed on the Exam 7029 eligibility list. The City appointed qualified candidates from the eligibility list in rank order. A candidate's rank/list number depended on (i) the candidate's scores on the written examination and PPT, (ii) the City's methodology for combining the two scores, (iii) the candidate's bonus points, and (iv) the City's methodology for breaking ties. The use of bonus points works to the advantage of African Americans and Hispanics and reduces the disparate impact of the rank-order process. Ties are broken in an essentially random process and should have no effect on the disparate impact of the rank-order process. Because the eligibility list for Exam 7029 was exhausted, I analyzed only the effect of the rank-order process in terms of delaying the appointment of African Americans and Hispanics.¹⁵

To determine whether the rank-order process resulted in a disparate impact on African Americans in terms of delay, I conducted a statistical test (*i.e.*, the Mann-Whitney Rank Sum Test) and found that African Americans were ranked statistically significantly lower than whites. The disparity in ranks is equivalent to 6.48 units of standard deviation. The likelihood that this large a disparity would occur by chance is less than 1-in-11 billion.

¹⁵ As noted earlier, however, being ranked lower on the eligibility list also affects a candidate's likelihood of being appointed, even if the candidate is reached on the eligibility list. I do not consider this effect.

-25-

In practical terms, the effect of the disparity in ranks was, on average, to delay the appointment of African Americans from the Exam 7029 eligibility list. Table 3 Part A presents the shortfall and surplus in African-American appointments by date of appointment, assuming that the appointment date distribution was statistically independent of race. Table 3 Part B shows the change in African-American appointment dates that would have had to occur to eliminate the disparate impact of the rank-order process. For example, Table 3 Part A shows that the expected number of African Americans appointed to the first academy class is 5, absent the disparity in ranks. Therefore, Table 3 Part B shows that the first 5 African Americans appointed all should have been appointed to the first class (*i.e.*, the February 4, 2001 class). Table 3 Part B also summarizes the total period of wages and seniority lost by the African-American firefighters whose appointment from the Exam 7029 eligibility list was delayed and the average loss for each African-American firefighter whose appointment was delayed. The 68 African-American firefighters whose appointments from Exam 7029 were delayed lost a total of approximately 20 years of wages and seniority as a result of the disparate impact of the rank-order process. The average loss for an individual African-American firefighter whose appointment was delayed was about 3 1/2 months of wages and seniority.

2. The City's Rank-Order Processing/Selection of Candidates from the Exam 7029 Eligibility List Resulted in a Disparate Impact upon Hispanics

To determine whether the rank-order process resulted in a disparate impact upon Hispanics, I conducted the same statistical test I conducted with regard to African Americans (*i.e.*, the Mann-Whitney Rank Sum Test) and found that Hispanics also were ranked statistically significantly lower than whites. The disparity in ranks is equivalent to 4.57 units of standard deviation. The likelihood that this large a disparity would occur by chance is less than 1-in-204,000.

-26-

In practical terms, the effect of the disparity in ranks was, on average, to delay the appointment of Hispanics from the Exam 7029 eligibility list. Table 4 Part A presents the shortfall and surplus in Hispanic appointments by date of appointment, assuming that the appointment date distribution was statistically independent of national origin. Table 4 Part B shows the change in Hispanic appointment dates that would have had to occur to eliminate the disparate impact of the rank-order process. Table 4 Part B also summarizes the total period of wages and seniority lost by Hispanic firefighters appointed from Exam 7029 and the average loss for each Hispanic firefighter whose appointment was delayed. The 86 Hispanic firefighters whose appointments from Exam 7029 were delayed due to the disparate impact of the rank-order process lost a total of approximately 23 years and 1 month of wages and seniority as a result of the disparate impact of the rank-order process. The average loss for an individual Hispanic firefighter whose appointment was delayed was slightly more than 3 months of wages and seniority.

C. Pass/Fail Use of Written Exam 2043

1. The City's Pass/Fail Use of Written Exam 2043 with a Cutoff Score of 70 Has Resulted in a Disparate Impact upon African-American Candidates

Table 5 shows that the City's pass/fail use of Written Exam 2043 has resulted in a statistically significant disparate impact upon African-American candidates. The disparity between the African-American pass rate and the white pass rate is equivalent to 21.84 units of standard deviation. The likelihood of such a disparity occurring by chance is infinitesimally small, less than 1-in-4.5 million billion.

To address the City's suggestion that this statistically significant finding is due to the large sample size, I again evaluated the results using a sample size reduced by 90% (so that there were

-27-

1,388 white candidates and 139 African-American candidates). If the same size disparity were to occur in the reduced sample, the disparity still would be statistically significant at 6.60 units of standard deviation. The likelihood of such a disparity occurring by chance is less than 1-in-2.4 billion. Again, the data does not support the City's suggestion.

In terms of practical effect, Table 5 shows that the City's pass/fail use of Written Exam 2043 with a cutoff score of 70 has resulted in a shortfall of 165 African-American written examination passers. That is, if African-American candidates had passed Written Exam 2043 at the same rate as white candidates, well over three-quarters (81.3%) of the African Americans who failed Written Exam 2043 would have passed it. Using the methodology described in Section IV. B. of my report, to date, an estimated 30 additional African Americans would have been appointed as firefighters.¹⁶ The ratio of the African-American and white pass rates is 87.8%.

2. The City's Pass/Fail Use of Written Exam 2043 with a Cutoff Score of 70 Has Resulted in a Disparate Impact upon Hispanic Candidates

Table 6 shows that the City's pass/fail use of Written Exam 2043 has resulted in a statistically significant disparate impact upon Hispanic candidates. The disparity between the Hispanic pass rate and the white pass rate is equivalent to 10.46 units of standard deviation. The likelihood of such a disparity occurring by chance is infinitesimally small, less than 1-in-4.5 million billion.

Again, to address the City's suggestion that this highly statistically significant finding is due to the large sample size, I evaluated the results using a sample size reduced by 90% (so that there were

¹⁶ Because the City continues to appoint firefighters from the Exam 2043 eligibility list, the number of appointments I estimate would have been made from the 165 additional African Americans who would have passed Written Exam 2043 had it not had a disparate impact will slightly increase as the City makes additional appointments, since the overall hire rate will increase.

-31-

between the Hispanic and white effective pass rates still would be statistically significant, equivalent to 3.24 units of standard deviation. The likelihood that such a disparity would occur by chance is less than 1-in-837.

In terms of practical effect, Table 9 shows that, as a result of the disparity between the Hispanic and white effective pass rates on Written Exam 2043, 242 Hispanics have been effectively eliminated from any possibility of appointment to date. Absent the disparity, to date, an estimated 45 additional Hispanics would have been appointed as FDNY firefighters. The ratio of the Hispanic effective pass rate to the white effective pass rate is 83.8%.²⁴

E. Rank-Order Processing/Selection of Candidates Based on a Combination of Written Exam 2043 and PPT Scores

1. The City's Use of Rank-Order Processing/Selection of Candidates Based on a Combination of Written Exam 2043 and PPT Scores Has Resulted in a Disparate Impact upon African-American Candidates

a. The City's Method of Ranking Candidates Based on a Combination of Written Exam 2043 and PPT Scores Has Resulted in Disproportionately Low Ranks for African-American Candidates

As explained previously, candidates who passed both Written Exam 2043 and the PPT were placed on the Exam 2043 eligibility list. The City appointed qualified candidates from the eligibility list in rank order (*i.e.*, in order by list number). A candidate's rank/list number depends on (i) the candidate's scores on the written examination and PPT; (ii) the City's methodology for combining

significant. As shown in Table 10, the disparity would be equivalent to 10.71 units of standard deviation. The likelihood that such a disparity would occur by chance is infinitesimally small, less than 1-in-4.5 million billion.

²⁴ Table 11 shows the comparable shortfall in Hispanic effective passers and the ratio of the Hispanic effective pass rate to the white effective pass rate if the City were to reach 400, 800 and 1,000 candidates further down the Exam 2043 before the list expires in May 2008.

-32-

the two scores; (iii) the candidate's bonus points; and (iv) the City's methodology for breaking ties. The use of bonus points works to the advantage of African Americans and Hispanics and reduces the disparate impact of the rank-order process. The City's handling of ties should have no effect on the disparate impact of the rank-order process because the tie-breaker (based on Social Security number) is essentially random.

The Mann-Whitney Rank Sum Test shows that African Americans are ranked statistically significantly lower than whites on the Exam 2043 eligibility list. The disparity between African-American and white ranks is equivalent to 9.45 units of standard deviation. The likelihood of such a disparity occurring by chance is infinitesimally small, less than 1-in-4.5 million billion.

b. Effect of Eliminating African-American Candidates from Possible Appointment

As stated previously, currently, the list number of the lowest-ranked candidate appointed from the Exam 2043 eligibility list is 4,825.5. In other words, to date, candidates with list numbers above 4,825.5 have had no possibility of being appointed. As explained in Section IV.D. of my report, such persons have "failed" the "ranking test," and we can measure the disparate impact of the City's rank-order process the same way we measure the impact of any pass/fail test.²⁵

Table 11 shows that the disparity in ranking "pass rates" between African Americans and whites is highly statistically significant (*i.e.*, the disparity between the percentages of African-

²⁵ Of course, this impact is not independent of the impact of the written examination, as measured by the effective cutoff score in the previous section of this report. The analysis summarized in this section of my report takes into consideration the effect of the City's method of assigning list numbers/ranks to candidates on the eligibility list. Thus, the current analysis takes into account candidates' performance on the written examination and PPT, the method the City has used to combine the written examination and PPT scores, the effect of bonus points and the City's method of handling ties.

-33-

American and white candidates who have list numbers of 4,825.5 or lower is statistically significant). The disparity is equivalent to 9.74 units of standard deviation. The likelihood of this disparity occurring by chance is infinitesimally small, less than 1-in-4.5 million billion.²⁶

In terms of practical effect, as a result of the disparity between the rates at which African Americans and whites pass the "ranking test," 95 African-American candidates have not ranked high enough to be appointed to date. An estimated 42 of the 95 would have been appointed as FDNY firefighters to date. The ratio of the African-American pass rate on the "ranking test" to the white pass rate is 67.6%.²⁷

c. Effect of Delay in Appointment of African-American Firefighters

Because appointments from among qualified candidates are made in rank order, on average, African-American candidates are reached on the list for possible appointment later than white candidates. As a result, the rank-order process also has delayed the appointment of some African-American candidates who have been appointed from Exam 2043. This delay effect is shown in Table 12. Table 12 Part B shows that, to date, the appointments of 44 African-American firefighters from Exam 2043 were delayed due to the disparate impact of the rank-order process. As a result, these 44

²⁶ If, before May 2008, when the Exam 2043 eligibility list expires, the City were to appoint candidates down to 1,000 candidates below the last candidate currently appointed, the disparity between the rates at which African Americans and whites would pass the "ranking test" would remain highly statistically significant. As shown in Table 11, the disparity would be equivalent to 9.78 units of standard deviation.

²⁷ Table 11 shows the comparable shortfalls and the ratio of the rate at which African Americans would pass the "ranking test" to the rate at which whites would pass the "ranking test" assuming the City were to reach 400, 800 and 1,000 candidates further down the Exam 2043 eligibility list before the list expired in May 2008. If the City were to reach 1,000 more candidates, the shortfall in African-American candidates reached would be 89, and the shortfall in appointments would be an estimated 37 African Americans. The ratio of the African-American and white pass rates would be 75.9%.

African-American firefighters have lost a total of approximately 14 years and 1 month of wages and seniority. Each of the 44 has lost, on average, approximately 4 months of wages and seniority.

2. The City's Rank-Order Processing/Selection of Candidates Based on a Combination of Written Exam 2043 and PPT Scores Has Resulted in a Disparate Impact upon Hispanic Candidates

a. The City's Method of Ranking Candidates Based on a Combination of Written Exam 2043 and PPT Scores Has Resulted in Disproportionately Low Ranks for Hispanic Candidates

The Mann-Whitney Rank Sum Test shows that Hispanics are ranked statistically significantly lower than whites on the Exam 2043 eligibility list. The disparity between Hispanic and white ranks is equivalent to 4.55 units of standard deviation. The likelihood of such a disparity occurring by chance is infinitesimally small, less than 1-in-186,225.

b. Effect of Eliminating Hispanic Candidates from Possible Appointment

Table 13 shows that the disparity in ranking "pass rates" between Hispanics and whites is statistically significant (*i.e.*, the disparity between the percentages of Hispanic and white candidates who have list numbers of 4,825.5 or lower is statistically significant). The disparity is equivalent to 5.04 units of standard deviation. The likelihood of this disparity occurring by chance is less than 1-in-2 million.²⁸

In terms of practical effect, as a result of the disparity between the rates at which Hispanics and whites pass the "ranking test," 63 Hispanic candidates have not ranked high enough to be

²⁸ If, before May 2008, when the Exam 2043 eligibility list expires, the City were to appoint candidates down to 1,000 candidates below the last candidate currently appointed, the disparity between the rates at which Hispanics and whites would pass the "ranking test" would remain statistically significant. As shown in Table 13, the disparity between the Hispanic and white pass rates would be equivalent to 4.03 units of standard deviation. The likelihood of this disparity occurring by chance is less than 1-in-17,920.

TABLE 1

**DISPARATE IMPACT
ON AFRICAN AMERICANS
OF PASS/FAIL USE OF WRITTEN EXAMINATION**

EXAM 7029

| | <u>WHITE</u> | <u>AFRICAN-AMERICAN</u> |
|---------------------------|--------------|-------------------------|
| Test Takers | 12,915 | 1,749 |
| Test Passers ¹ | 11,613 | 1,054 |
| Pass Rate | 89.92% | 60.26% |

Difference in pass rate by race:

In units of standard deviation² 33.9

Likelihood of as large a disparity occurring by chance 1-in-4.5 million billion

Shortfall in African-American test passers 519

Estimated shortfall in African-American appointments 114

Ratio of African-American pass rate to white pass rate
(80% Rule) 67.0%

NOTES

¹ Scoring 84.705 or higher on written examination

² Based on Yates Corrected Chi-Square calculation

TABLE 3
PART A**IMPACT OF RANKING
ON TIME TO APPOINTMENT OF AFRICAN AMERICANS****EXAM 7029**

| <u>APPOINTMENT DATE</u> | <u>AFRICAN-AMERICAN APPOINTMENTS</u> | | <u>SHORTFALL³</u> |
|-----------------------------|--|-----------------------------|------------------------------|
| | <u>ACTUAL¹</u> | <u>EXPECTED²</u> | |
| 02/04/01 | 2 | 5 | 3 |
| 05/06/01 | 1 | 3 | 2 |
| 07/15/01 | 1 | 5 | 4 |
| 10/28/01 | 10 | 10 | 0 |
| 01/27/02 | 12 | 9 | -3 |
| 05/06/02 | 8 | 10 | 2 |
| 07/28/02 | 14 | 9 | -5 |
| 02/02/03 | 6 | 9 | 3 |
| 05/04/03 | 10 | 11 | 1 |
| 09/14/03 | 13 | 7 | -6 |
| 12/10/03 | 6 | 9 | 3 |
| 03/07/04 | 4 | 8 | 4 |
| 05/25/04 | 8 | 5 | -3 |
| 9/12/2004 ⁴ | 9 | 4 | -5 |
| TOTAL | 104 | 104 | 0 |

NOTES

- ¹ Based on actual appointment dates
- ² If appointment dates were statistically independent of race, rounded to whole person
- ³ Negative shortfall means surplus of African-American appointments
- ⁴ Appointment date 9/12/2004 means on or after that date

TABLE 3
PART B

**IMPACT OF RANKING
ON TIME TO APPOINTMENT OF AFRICAN AMERICANS**

EXAM 7029

| <u>APPOINTMENT DATE</u> | <u>NUMBER OF AFRICAN- AMERICAN APPOINTMENTS¹</u> | <u>DATE SHOULD HAVE BEEN APPOINTED</u> | <u>LOST YEARS OF WAGES/SENIORITY PER PERSON</u> | <u>TOTAL LOST YEARS OF WAGES/SENIORITY²</u> |
|-----------------------------|---|--|---|--|
| 02/04/01 | 2 | 02/04/01 | 0.00 | 0.00 |
| 05/06/01 | 1 | 02/04/01 | 0.25 | 0.25 |
| 07/15/01 | 1 | 02/04/01 | 0.44 | 0.44 |
| 10/28/01 | 1 | 02/04/01 | 0.73 | 0.73 |
| 10/28/01 | 3 | 05/06/01 | 0.48 | 1.44 |
| 10/28/01 | 5 | 07/15/01 | 0.29 | 1.44 |
| 10/28/01 | 1 | 10/28/01 | 0.00 | 0.00 |
| 01/27/02 | 9 | 10/28/01 | 0.25 | 2.24 |
| 01/27/02 | 3 | 01/27/02 | 0.00 | 0.00 |
| 05/06/02 | 6 | 01/27/02 | 0.27 | 1.63 |
| 05/06/02 | 2 | 05/06/02 | 0.00 | 0.00 |
| 07/28/02 | 8 | 05/06/02 | 0.23 | 1.82 |
| 07/28/02 | 6 | 07/28/02 | 0.00 | 0.00 |
| 02/02/03 | 3 | 07/28/02 | 0.52 | 1.55 |
| 02/02/03 | 3 | 02/02/03 | 0.00 | 0.00 |
| 05/04/03 | 6 | 02/02/03 | 0.25 | 1.50 |
| 05/04/03 | 4 | 05/04/03 | 0.00 | 0.00 |
| 09/14/03 | 7 | 05/04/03 | 0.36 | 2.55 |
| 09/14/03 | 6 | 09/14/03 | 0.00 | 0.00 |
| 12/10/03 | 1 | 09/14/03 | 0.24 | 0.24 |
| 12/10/03 | 5 | 12/10/03 | 0.00 | 0.00 |
| 03/07/04 | 4 | 12/10/03 | 0.24 | 0.96 |
| 05/25/04 | 8 | 03/07/04 | 0.22 | 1.73 |
| 09/12/04 | 5 | 05/25/04 | 0.30 | 1.51 |
| 09/12/04 | 4 | 09/12/04 | 0.00 | 0.00 |
| TOTAL | 104 | | | 20.03 |
| DELAYED | 68 | | | |

**AVERAGE LOST YEARS OF WAGES/SENIORITY
PER DELAYED AFRICAN AMERICAN**

0.29 (3.48 MONTHS)

NOTES

¹ Based on actual appointment dates

² Total lost years of wages/seniority (which is the product of the number of African-American appointments times the lost years of wages/seniority per person) may not appear equal due to rounding

TABLE 5

**DISPARATE IMPACT
ON AFRICAN AMERICANS
OF PASS/FAIL USE OF WRITTEN EXAMINATION**

EXAM 2043

| | <u>WHITE</u> | <u>AFRICAN-AMERICAN</u> |
|---------------------------|--------------|-------------------------|
| Test Takers | 13,878 | 1,393 |
| Test Passers ¹ | 13,496 | 1,190 |
| Pass Rate | 97.25% | 85.43% |

Difference in pass rate by race:

In units of standard deviation²

21.84

Likelihood of as large a disparity occurring by chance

1-in-4.5 million billion

Shortfall in African-American test passers

165

Current estimated shortfall in African-American appointments

30

Ratio of African-American pass rate to white pass rate
(80% Rule)

87.8%

NOTES

¹ Scoring 70.588 or higher on written examination

² Based on Yates Corrected Chi-Square calculation

TABLE 11

IMPACT OF RANKING ON APPOINTMENT OF AFRICAN AMERICANS

EXAM 2043

| NUMBER OF CANDIDATES REACHED ON LIST | "RANKING TEST" PASS RATE¹ | | DISPARITY IN UNITS OF STANDARD DEVIATION | RATIO OF PASS RATES (80% RULE) | SHORTFALL IN | |
|---|---|----------------------|---|--------------------------------------|--|---|
| | WHITE | AFRICAN- AMERICAN | | | REACHED FOR POSSIBLE APPOINTMENT | ESTIMATED HIRES DUE TO RANKING |
| CURRENT: 4862 | 68.68% | 46.41% | 9.74 | 67.6% | 95 | 42 |
| (400 more): 5262 | 72.62% | 51.20% | 9.71 | 70.5% | 91 | 40 |
| (800 more): 5662 | 78.16% | 58.17% | 9.72 | 74.4% | 85 | 37 |
| (1,000 more): 5862 | 80.40% | 61.00% | 9.78 | 75.9% | 83 | 30 |

NOTE

¹ Ranking high enough on eligibility list to be reached for possible appointment

TABLE 12
PART A

**IMPACT OF RANKING
ON TIME TO APPOINTMENT OF AFRICAN AMERICANS**

EXAM 2043

| <u>APPOINTMENT DATE</u> | <u>AFRICAN-AMERICAN APPOINTMENTS</u> | | <u>SHORTFALL³</u> |
|-----------------------------|--|-----------------------------|------------------------------|
| | <u>ACTUAL¹</u> | <u>EXPECTED²</u> | |
| 05/25/04 | 0 | 1 | 1 |
| 09/12/04 | 0 | 2 | 2 |
| 12/05/04 | 2 | 6 | 4 |
| 03/08/05 | 12 | 9 | -3 |
| 05/31/05 | 3 | 9 | 6 |
| 09/25/05 | 12 | 6 | -6 |
| 01/15/06 | 3 | 4 | 1 |
| 04/11/06 | 8 | 6 | -2 |
| 06/11/06 | 5 | 7 | 2 |
| 11/19/06 | 7 | 8 | 1 |
| 03/25/07 | 17 | 11 | -6 |
| 08/05/07 | 11 | 11 | 0 |
| TOTAL | 80 | 80 | 0 |

NOTE

¹ Based on actual appointment dates

² If appointment dates were statistically independent of race, rounded to whole person

³ Negative shortfall means surplus of African-American appointments

TABLE 12
PART B

**IMPACT OF RANKING
ON TIME TO APPOINTMENT OF AFRICAN AMERICANS**

EXAM 2043

| APPOINTMENT DATE | NUMBER OF AFRICAN- AMERICAN APPOINTMENTS¹ | DATE SHOULD HAVE BEEN APPOINTED | LOST YEARS OF WAGES/SENIORITY PER PERSON | TOTAL LOST YEARS OF WAGES/SENIORITY² |
|-----------------------------|---|--|---|--|
| 12/05/04 | 1 | 05/25/04 | 0.53 | 0.53 |
| 12/05/04 | 1 | 09/12/04 | 0.23 | 0.23 |
| 03/08/05 | 1 | 09/12/04 | 0.48 | 0.48 |
| 03/08/05 | 6 | 12/05/04 | 0.25 | 1.53 |
| 03/08/05 | 5 | 03/08/05 | 0.00 | 0.00 |
| 05/31/05 | 3 | 03/08/05 | 0.23 | 0.69 |
| 09/25/05 | 1 | 03/08/05 | 0.55 | 0.55 |
| 09/25/05 | 9 | 05/31/05 | 0.32 | 2.88 |
| 09/25/05 | 2 | 09/25/05 | 0.00 | 0.00 |
| 01/15/06 | 3 | 09/25/05 | 0.31 | 0.92 |
| 04/11/06 | 1 | 09/25/05 | 0.54 | 0.54 |
| 04/11/06 | 4 | 01/15/06 | 0.24 | 0.94 |
| 04/11/06 | 3 | 04/11/06 | 0.00 | 0.00 |
| 06/11/06 | 3 | 04/11/06 | 0.17 | 0.50 |
| 06/11/06 | 2 | 06/11/06 | 0.00 | 0.00 |
| 11/19/06 | 5 | 06/11/06 | 0.44 | 2.21 |
| 11/19/06 | 2 | 11/19/06 | 0.00 | 0.00 |
| 03/25/07 | 6 | 11/19/06 | 0.35 | 2.07 |
| 03/25/07 | 11 | 03/25/07 | 0.00 | 0.00 |
| 08/05/07 | 11 | 08/05/07 | 0.00 | 0.00 |
| TOTAL | 80 | | | 14.08 |
| DELAYED | 44 | | | |

**AVERAGE LOST YEARS OF WAGES/SENIORITY PER
DELAYED AFRICAN AMERICAN**

0.32 (3.84 MONTHS)

NOTES

¹ Based on actual appointment dates

² Total lost years of wages/seniority (which is the product of the number of African-American appointments times the lost years of wages/seniority per person) may not appear equal due to rounding

EXHIBIT P

1

1

2 UNITED STATES DISTRICT COURT

3 EASTERN DISTRICT OF NEW YORK

4 CIV. ACTION NO.: 07-CV-2067

5 (GARAUFIS, J.)(MANN, M.J.)

6 -----x

7 UNITED STATES OF AMERICA,

8 Plaintiff,

9 VULCAN SOCIETY, INC., FOR ITSELF AND ON
BEHALF OF ITS MEMBERS, CANDIDO NUNEZ,
10 MARCUS HAYWOOD AND ROGER GREGG,
INDIVIDUALLY AND ON BEHALF OF A CLASS OF
11 ALL OTHERS SIMILARLY SITUATED,

12 Plaintiff-Intervenors,
-against-

13

CITY OF NEW YORK, FIRE DEPARTMENT OF THE
14 CITY OF NEW YORK, NEW YORK CITY DEPARTMENT
OF CITYWIDE ADMINISTRATIVE SERVICES, MAYOR
15 MICHAEL BLOOMBERG AND NEW YORK CITY FIRE
COMMISSIONER NICHOLAS SCOPPETTA, IN THEIR
16 INDIVIDUAL AND OFFICIAL CAPACITIES,

17 Defendants.

18 -----x

19 100 Church Street
New York, New York

20

October 9, 2007
21 10:15 a.m.

22

23 DEPOSITION of THOMAS PATITUCCI

24

25

1 T. Patitucci
2 was not why the pass mark for Written Exam
3 2043 was set at 70?

4 MR. SAMPLE: Objection.
5 Repetitive.

6 Q Why was the pass mark for
7 Written Exam 2043 set at 70?

8 A Because that is the default pass
9 mark in our personnel rules and
10 regulations.

11 Q And so you are representing to
12 me today that the setting of the pass mark
13 for Written Exam 2043 was not done because
14 of the FDNY's hiring needs?

15 MR. SAMPLE: Objection.
16 Repetitive.

17 A That's correct.

18 Q If you will look at the second
19 paragraph on the page Bates numbered
20 DCAS-E-00006612 at the first sentence it
21 says: "Administration of the previous
22 firefighter multiple-choice test, Exam No.
23 7029, resulted in adverse impact against
24 Black candidates."

25 Is that what it says?

1 T. Patitucci

2 lowering the FDNY standards because in the
3 past, the FDNY has sent an unrealistically
4 pass mark on our exam."

5 Is that what it says?

6 A Yes.

7 Q What did you mean when you said
8 that "in the past, the FDNY has set an
9 unrealistically pass mark on our exam"?

10 A I felt that FDNY was setting the
11 pass mark and they were high.

12 Q Unrealistically high?

13 A Unrealistically high.

14 Q Was that with respect to Written
15 Exam 7029?

16 A Yes.

17 Q And so the pass mark of Written
18 Exam 7029 was 84.705; is that correct?

19 A That's correct.

20 Q And it is your contention that
21 that was unrealistically high?

22 MR. SAMPLE: Objection.

23 Repetitive.

24 A Yes.

25 Q What was the basis for your

1 T. Patitucci

2 belief that the pass mark for Written Exam
3 7029 was unrealistically high?

4 A Primarily because our rules and
5 regulations say the pass mark should be
6 70.

7 Q Do you recall protesting to
8 anyone within DCAS about the setting of
9 the pass mark for Written Exam 7029?

10 A I objected to the fact that the
11 fire department was setting the pass mark.

12 Q To whom did you object?

13 A Carol Wachter.

14 Q Did you object to anyone else?

15 A No.

16 Q What was Ms. Wachter's response
17 to your objections?

18 A That we were going to do it any
19 way. I guess she had -- well, basically
20 she said we have to do it any way.

21 Q DCAS had no choice in setting
22 the pass mark?

23 A I don't think she had a choice.

24 Q Mr. DeMarco was Ms. Wachter's
25 superior; is that correct?

1 T. Patitucci

2 A That's correct.

3 Q Would Mr. DeMarco have had a
4 choice?

5 A I really can't answer that
6 question, because I don't know where the
7 influence was coming from, who was making
8 the decision that FDNY should set the pass
9 mark.

10 Q Who else could have made the
11 decision?

12 A The Commissioner.

13 Q Of DCAS?

14 A The Commissioner of DCAS at the
15 time.

16 Q The next document in the stack
17 immediately to your right should be what
18 has been previously marked as Exhibit 302.

19 Mr. Patitucci, can you identify
20 this document?

21 A (Reviewing exhibit.) Again,
22 really, it doesn't look familiar.

23 Q I am going to represent to you
24 that this has been previously identified
25 as a memorandum from Nicholas Scoppetta to

91

1 T. PATITUCCI

2 states the fire department has requested
3 that a total of 15,000 candidates from
4 both this test and the promotion test be
5 called to the physical. At this pass mark
6 the fire department's request is met. Is
7 this an accurate statement of the reason
8 that the passing standard on Written Exam
9 7029 was set at 84.75?

10 A. I believe we mentioned before I
11 believe it was set based on agency need,
12 yes.

13 Q. That was the request of the fire
14 department; is that correct?

15 A. Yes.

16 Q. Do you know why the fire
17 department wanted 15,000 candidates from
18 this test and the promotional test to be
19 called to the physical?

20 A. Again it was based on agency
21 need, so you would have to ask them how
22 they worked out that formula to get to
23 that point.

24 Q. Did you ever discuss the cutoff
25 score with Mr. Morrongiello on this

92

1 T. PATITUCCI

2 Written Exam 7029?

3 A. Basically I probably told him

4 that's where he should set it because

5 that's what I was told. Basically we were

6 told to set the pass performance at 74.75.

7 We were provided documentation or test

8 results from the fire department and on

9 the basis of that and I assume their

10 agency need, they looked at how people

11 performed on the test and agency need and

12 they told us that that was the pass mark

13 they wanted.

14 Q. In setting the passing score on

15 Written Exam 7029 then, was input sought

16 from anyone outside the fire department

17 with regard to the cutoff?

18 A. Not that I'm aware of unless the

19 fire department did. You know, we

20 basically defer, as I said, to the fire

21 department on this issue. If they spoke

22 to other people about it, then they may

23 have, I don't know. But we did not.

24 Q. Now, was a job analysis done in

25 connection with Written Exam 7029?

93

1 T. PATITUCCI

2 A. Yes, it was.

3 Q. Then is it correct that the
4 cutoff score on Written Exam 7029 wasn't
5 based in any way on any of the data or
6 information that was gathered as part of
7 the job analysis?

8 A. The cutoff score was based upon
9 apparently agency need based upon the
10 document that you see in front of you
11 right now, Exhibit 25.

12 Q. As far as you know, there was no
13 other reason for selecting that cutoff
14 score; is that correct?

15 A. As far as I know, no.

16 MS. SEELEY: Mark this as
17 Exhibit 26, please.

18 (Whereupon, the above-mentioned
19 document was marked Plaintiffs'
20 Exhibit 26 for identification.)

21 Q. Mr. Patitucci, I'm handing you
22 what's been marked as Exhibit 26. Again,
23 this does not have USA Bates number, but
24 it's Bates numbered in the lower right
25 DCAS 004504507. Have you seen these

1 T. Patitucci

2 mean, this may be in response to something
3 that she said down here.

4 I don't -- I don't recall what I
5 meant by that.

6 Q Mr. Patitucci, is it true that
7 DCAS never calculated the reliability of
8 the PPT?

9 A That's correct.

10 Q Do you know whether the PPT had
11 adequate reliability for use?

12 A I assume the -- no, I don't.

13 Q If an exam is not reliable, how
14 can it be valid?

15 A It cannot be.

16 Q Mr. Patitucci, I have one more
17 question I want to ask about this exhibit.

18 A Okay.

19 Q The first sentence on
20 Exhibit 509, you say: "I will call you to
21 discuss how the pass mark was set for the
22 overall exam."

23 What were you referring to by
24 the pass mark set for the overall exam?

25 A I guess it depends on -- I don't

1 T. PATITUCCI

2 validated for use in New York City back

3 when it was developed.

4 Q. I guess I'm trying to understand

5 when you talk about its accuracy in

6 selecting people, in fact, to the present

7 because you're still hiring from Exam

8 2043, right?

9 A. Yes.

10 Q. The people who went through the

11 academy and then took the CPAT at the end

12 for the state, they had already been

13 screened with the PPT, right?

14 A. Yes.

15 Q. And they had already been ranked

16 with the PPT and the department?

17 A. That's right.

18 Q. How do you know that the CPAT

19 was accurately selecting people?

20 A. Well, the fire department was

21 telling us that there was a very strong

22 relationship between passing the test and

23 being able to perform the duties of a

24 firefighter. Plus they showed us

25 information from the rest of the test

EXHIBIT Q

1

2 IN THE UNITED STATES DISTRICT COURT
3 FOR THE EASTERN DISTRICT OF NEW YORK

4 Civil Action No. 07-CV-2067

-----x

5 UNITED STATES OF AMERICA,

6 PLAINTIFF,

7 AND

8 VULCAN SOCIETY, INC., FOR ITSELF AND ON
9 BEHALF OF ITS MEMBERS, CANDIDO NUNEZ,
10 MARCUS HAYWOOD AND ON BEHALF OF A CLASS
11 OF ALL OTHERS SIMILARLY SITUATED,

12

13 PLAINTIFF-INTERVENORS,

14

15 V.

16

17 CITY OF NEW YORK, FIRE DEPARTMENT OF THE
18 CITY OF NEW YORK, NEW YORK CITY
19 DEPARTMENT OF CITYWIDE ADMINISTRATIVE
20 SERVICES, MAYOR MICHAEL BLOOMBERG AND NEW
21 YORK CITY FIRE COMMISSIONER NICHOLAS
22 SCOPPETTA, IN THEIR INDIVIDUAL AND
23 OFFICIAL CAPACITIES,

24

25 DEFENDANTS.

-----x

October 23, 2007

10:24 a.m.

DEPOSITION of STEPHEN DOBROWSKY, taken

by Plaintiff, pursuant to Rule 30(b)(6)

Notice, held at the New York City Law

Department, 100 Church Street, New York,

before Jamie Ann Stanton, a Shorthand

Reporter and Notary Public of the State

of New York

* * *

59

1 S. Dobrowsky

2 portions of it that deal with

3 Certifications?

4 A I would hope I would be.

5 Q Let me ask you about some of the

6 specific parts of the rule.

7 A Okay.

8 Q If you go to the page that's

9 Bates numbered on the lower left USA

10 006267. And, in particular, Rule

11 4.7.1(c).

12 Do you see that?

13 A Okay. Yes.

14 Q Do you want to take a minute to

15 look at that?

16 A (Complying.) Okay.

17 Q I want to try to make sure that

18 I understand the process of certifying

19 people from the firefighter eligible list.

20 So 4.7.1(c) seems to indicate that

21 candidates will be certified by DCAS three

22 at a time.

23 First, in the context of this

24 rule, what does certified mean?

25 A Certified is a term that we use

60

1 S. Dobrowsky

2 when we send a list of names to an agency

3 for permanent appointment purposes.

4 Q So in this case?

5 A So a certified list we generally

6 refer to as that certification list.

7 Q The City appoints entry-level

8 firefighters essentially a class at a

9 time, right?

10 A That's correct.

11 Q And how big is a class, usually,

12 for firefighters?

13 A It varies. About 150 seems to

14 be a common class size.

15 Q So the City hires firefighters

16 around 150 or so at a time?

17 A That's a number that comes up

18 quite often as a class.

19 Q So how many candidates from a

20 firefighter eligibility list are certified

21 at a time?

22 A It varies. I would have to

23 check. We would send out several hundred

24 names, possibly a thousand names to the

25 agency. Depending at the stage of the

EXHIBIT R

g. Other Statistical Tests

The Chi Square test is widely used with categorical data. Other statistical tests are needed for other situations, such as comparing average scores of two groups. To compare the numeric averages of two groups, either the F test or the t-test is appropriate. I indicate in the tables below when I use either of these tests, as is professionally appropriate in the reporting of such analyses. Both of these tests involve interpreting a test statistic calculated from the data. Both tests use degrees of freedom to help find the probability associated with the calculated value of the test statistic. This probability is the probability that an outcome as extreme as the observed outcome would occur by chance alone.

h. What is Statistical Significance?

In the coin toss example we said that a probability of 1 in 500 (or .002) is too rare to accept as happening purely by chance, and so we decided that the coin was biased. The most usual standard of statistical proof uses .05 as the critical value. If the probability of an outcome as extreme as the observed outcome (or result) is .05 or less, we say the result is statistically significant and we say that is improbable that the outcome occurred due to chance alone. In many commonly used statistical tests, a probability of .05 corresponds to 1.96 standard deviation units. Perhaps the next most commonly used criterion is .01, which corresponds to 2.56 standard deviation units. This is the basis for the US Supreme Court's reference to "two or three standard deviations" as a general rule for evaluating statistical significance (*Castaneda v. Partida*, 430 U.S. 482, 496-497, n. 17).

I. What is the difference Between a One and Two Tailed Test?

A statistical test asks the question, what is the probability of an outcome as unusual as we have observed. Often when we have two possible outcomes we are concerned about differences in both directions (e.g., biased in favor of heads and biased in favor of tails). Sometimes when we have two possible outcomes we are interested in identifying an unusual outcome in one direction. For example, if we have an established medical procedure, we may be interested in deciding if a new procedure is better than the existing method, and we may have no interest in deciding if it is worse. When we are interested in identifying unusual events in either direction, we use a two-tailed test. When we are interested in identifying events in one direction, we use a one-tailed test. Usually experts for plaintiffs like to use one-tailed tests, since it is easier to achieve statistical significance with one tailed tests, and because the usual employment discrimination case involves adverse impact against minorities, and ignores the possibility of bias in favor of minorities. Experts for the defense usually prefer two tailed tests, for the same reason. In the analyses I conducted for this report, I use the conservative, two-tailed approach unless otherwise noted.

Section Two: Adverse Impact Analyses: 1999 Exam

1. Adverse Impact in Appointments by Race: Exam 7029 (1999)

The most basic evaluation of adverse impact involves an evaluation of the number and percent of black and white applicants appointed (or hired). I did an adverse impact analysis of appointments from the 1999 exam, as follows.

a. Number of Appointments: Exam 7029 (1999)

I tallied the number of persons appointed based on data in an Excel file provided to me titled "exam7029 corrected applicant data.xls - (# Legal 1820012).XLS". For the sake of this analysis, and other similar analyses below, I included as appointed those applicants with first or last disposition action codes that indicated they were: appointed, appointed from a promotional list, offered a job and turned it down, or failed to report after accepting appointment (disposition action codes APP, AOL, DEA, and FRA).⁵ For the sake of this analysis, and other similar analyses below, I excluded those applicants with disposition action codes that indicate they could not be appointed for reasons that were generally of a voluntary nature. The excluded disposition codes indicated that an applicant: was deceased, failed to report for investigation, failed to report for interview, was not qualified for appointment, was overage for the position, or was underage at time of appointment (disposition action codes DCE, FRI, FTR, NQA, OVA, UNA). These exclusions were based on a combination of the first and last disposition code: the first disposition code had to be one of those just listed, and the last disposition code had to be either one of those just listed or blank (indicating the applicant was not certified again).⁶

These tallies are presented in Table 1a below, with the associated percents. This table shows 6.7% of the black applicants were appointed, as compared with 23.6% of the white applicants.

⁵The data file available to me contained only disposition action codes for the first and last certifications. Although these are arguably the most important certifications, if and when I get disposition code information for any other certifications, I may redo this analysis.

⁶I replicated this analysis counting only appointments (disposition code APP), and without the exclusions just noted, and the results were virtually unchanged as to adverse impact and statistical significance. This analysis is summarized in Attachment D.

There is practically important and highly statistically significant adverse impact against black applicants in appointment of males.

3. Adverse Impact in Passing the Written Test: Exam 7029 (1999)

There are many possible reasons for the adverse impact just shown in the appointment of black firefighters, ranging from failing the examination, to low scores on the written test and PPT that comprise the examination, to low placement on the eligible list, to differential impact of the City's post-exam processing and screening, to applicants' decisions to drop out of contention. In the next several analyses, we will go through the examination process basically in chronological order of its administration and the subsequent use of the eligible list. To pass the examination an applicant must pass both the written test (with a score of 84.705) and pass the PPT. I did an adverse impact analysis of the pass-fail decisions for the written test of the 1999 exam.¹²

a. Number Passing and Failing the Written Test: Exam 7029 (1999)

I tallied the number of applicants passing and failing the written test alone based on data in an Excel file provided to me titled "exam7029 corrected applicant data.xls - (# Legal 1820012).XLS". These tallies are presented in Table 3a below, with the associated percents.

| Table 3a. Pass-Fail Written Test of Exam 7029, by Racial Group | | | |
|---|--|--|----------------------------|
| Group | Number Pass Written (percent) | Number Fail Written (percent) | Total (percent) |
| Black | 1,054 (60.3%) | 695 (39.7%) | 1,749 (100%) |
| White | 11,613 (89.9%) | 1,302 (10.1%) | 12,915 (100%) |
| Total | 12,667 | 1,997 | 14,664 |

b. Adverse Impact Ratio for Passing Written Test: Exam 7029 (1999)

As shown in Table 3b, the adverse impact ratio for passing the written test is .67, which I got by dividing the percentage of blacks passing the written test (60.3%) by the percentage of whites passing (89.9%). This result tells us that black applicants passed the written test at a rate of 67% of the rate that white applicants did, or that black

¹²Sometimes the pass-fail point has little practical import, as in cases where there are so many applicants who passed that only high scoring applicants are considered for appointment. That was not the case for the 1999 exam. For the 1999 exam, 14 applicants with the lowest passing score on the written test, 84.705, were appointed.

applicants passed the written test at about 2/3 the rate of white applicants. A statistical analysis of this adverse impact ratio is reported in Table 3b.

| Table 3b. Adverse Impact: Pass-Fail Written Test of Exam 7029 | | | | | | |
|---|---------------------------|----------------------------|-------------|------------|--------------------------------|-----------|
| Black: Pass Written | White: Pass Written | Adverse Impact Ratio | Probability | Chi Square | Standard Deviation Units | Shortfall |
| 60.3% | 89.9% | .67 | < .0001 | 1149.1 | 33.9 | 457 |

- c. **Probability and Statistical Significance of Passing Written Test: Exam 7029 (1999)**
The adverse impact ratio of .67 is highly statistically significant. A ratio this small would occur by chance much less than one time in ten thousand (see Table 3b, the column titled Probability). This probability corresponds to 33.9 standard deviation units.
 - d. **Practical Import of Passing Written Test: Exam 7029 (1999)**
The practical importance of 60.3% of black applicants passing the written test versus 89.9% of white applicants is obvious: there is disproportionate passing of this hurdle, with a shortfall of 457 black passers. This adverse impact ratio (.67) also fails the .80 rule of thumb of the *UGESP*. The City passed white applicants on the written test at about 1½ times the rate of black applicants.
 - e. **Opinion: Pass-Fail Written Test: Exam 7029 (1999)**
There is practically important and highly statistically significant adverse impact against black applicants in passing the written test.
4. **Adverse Impact in Passing the Physical Performance Test: Exam 7029 (1999)**
In addition to passing the written test, an applicant must pass the Physical Performance Test PPT (with a score of 75) to pass the exam. I did an adverse impact analysis of the pass-fail decisions¹³ for the PPT of the 1999 exam, as follows.
 - a. **Number Passing and Failing the PPT: Exam 7029 (1999)**
I tallied the number of applicants passing and failing the PPT alone based on data in an Excel file provided to me titled "exam7029 corrected applicant data.xls - (# Legal 1820012).XLS". Here, and in subsequent analyses of PPT scores, I assumed an applicant took the PPT if he or she had a non-zero entry in the field "Date of Physical Test" or had a raw PPT score other than zero. The tallies are presented in Table 4a

¹³Sometimes the pass-fail point has little practical import, as in cases where there are so many applicants who pass that only high scoring applicants are considered for appointment. That was not the case for the 1999 PPT: 129 applicants with the lowest passing score on the PPT were appointed.

black applicants in the average scores on the written test and PPT.

9. Adverse Impact in Rank Order Placement On the List: Exam 7029 (1999)

Focusing on the passing applicants, we can ask if there was adverse impact in the ranking of black applicants who passed the exam. This is important because no one can be appointed before they are certified, and certification is in rank order starting with the applicant ranked first on the list (who is the applicant with the highest score). Therefore, I did an adverse impact analysis of the rank order placement, comparing the placement of black and white applicants in two ways: average rank, and proportion in each grouping of 1,000 ranks. I used groupings of 1,000 ranks because the NYC FD has hired, on average, about 800 Firefighters yearly since 2002, the first year that Firefighters were hired from the 1999 exam. My adverse impact analysis of rank order placement on the list follows.¹⁶

a. Average Rank on the List: Exam 7029 (1999)

I calculated the average (i.e., the mean) rank on the list for black and white applicants based on an Excel file provided to me titled "7029 Spreadsheet FINAL.xls". The average ranks on the list are presented in Table 9a below.

| Table 9a. Average Rank on the List, by Race | |
|---|------------|
| Group | Mean Rank* |
| Black | 3,813 |
| White | 3,183 |

* Lower numbered ranks are certified before higher numbered ranks.

b. Probability and Statistical Significance of Average Rank on the List: Exam 7029

The difference between the average rank on the list for black and for white applicants is 630 ranks, and this difference is highly statistically significant (see Table 9b). A difference this large would occur by chance less than one time in ten thousand. This corresponds to 6.5 standard deviation units.

| Table 9b. Average Rank on the List, by Race: Statistical Significance | | | | | |
|---|--------------------|------------|-------------|--------------|--------------------------|
| Black Average Rank | White Average Rank | Difference | Probability | F Statistic* | Standard Deviation Units |
| 3,813 | 3,183 | 630 | < .0001 | 42.6 | 6.5 |

* With 1 and 5,660 d.f.

¹⁶I replicated this analysis for males only, and the results were virtually unchanged as to adverse impact and statistical significance.

2/5 to 2/3 the rate of white applicants, depending on the passmark. All these adverse impact ratios fail the .80 rule of thumb of the *UGESP*. In addition, it is clear that using this test to rank applicants will result in additional adverse impact among passing applicants, due to ranking.

d. Opinion on Written Test Adverse Impact Ratios from Exhibit 26: Exam 7029 (1999)

There is practically important and highly statistically significant adverse impact against black applicants in passing the written test at each of the ten passmarks reported by the City in Exhibit 26.

Section Three: Adverse Impact Analyses: 2002 Exam

14. Adverse Impact in Appointments by Race: Exam 2043 (2002)

The most basic evaluation of adverse impact involves an evaluation of the number and percent of black and white applicants appointed (or hired). Although appointments may still be made from this exam, I did an adverse impact analysis of the appointments resulting from the 2002 exam as of the effective date of the data file provided by the City, as follows.

a. Number of Appointments: Exam 2043 (2002)

I tallied the number of persons appointed based on data in an Excel file provided to me titled "2043 Spreadsheet FINAL.xls". For the sake of this analysis, and other similar analyses below, I included as appointed those applicants with first or last disposition action codes that indicated they were: appointed, appointed from a promotional list, offered a job and turned it down, or failed to report after accepting appointment (disposition action codes APP, AOL, DEA, and FRA).²¹ For the sake of this analysis, and other similar analyses below, I excluded those applicants with disposition action codes that indicate they could not be appointed for reasons that were generally of a voluntary nature. The excluded disposition codes indicated that an applicant: was deceased, failed to report for investigation, failed to report for interview, was not qualified for appointment, was overage for the position, or was underage at time of appointment (disposition action codes DCE, FRI, FTR, NQA, OVA, UNA). These exclusions were based on a combination of the first and last disposition code: the first disposition code had to be one of those just listed, and the last disposition code had to be either one of those just listed or blank (indicating the applicant was not certified again).²²

²¹The data file available to me contained only disposition action codes for the first and last certifications. Although these are arguably the most important certifications, if and when I get disposition code information for any other certifications, I may redo this analysis.

²²I replicated this analysis counting only appointments (disposition code APP), and without the exclusions just noted, and the results were virtually unchanged as to adverse impact

written test of the 2002 exam.²³

a. Number Passing and Failing the Written Test: Exam 2043 (2002)

I tallied the number of applicants passing and failing the written test alone based on data in an Excel file provided to me titled "2043 Spreadsheet FINAL.xls". These tallies are presented in Table 16a below, with the associated percents.

| Table 16a. Pass-Fail Written Test of Exam 2043, by Racial Group | | | |
|--|--|--|----------------------------|
| Group | Number Pass Written (percent) | Number Fail Written (percent) | Total (percent) |
| Black | 1,190 (85.4%) | 203 (14.6%) | 1,393 (100%) |
| White | 13,495 (97.2%) | 382 (2.8%) | 13,877 (100%) |
| Total | 14,685 | 585 | 15,270 |

b. Adverse Impact Ratio for Passing Written Test: Exam 2043 (2002)

As shown in Table 16b, the adverse impact ratio for passing the written test is .88, which I got by dividing the percentage of blacks passing the written test (85.4%) by the percentage of whites passing (97.2%). This result tells us that black applicants passed the written test at a rate of 88% of the rate that white applicants pass. Stated differently, less than 3% of the white applicants failed the written test, while 14.6% of the black applicants failed the test, more than 5 times the failure rate of white applicants. A statistical analysis of this adverse impact ratio is reported in Table 16b.

²³Sometimes the pass-fail point has little practical import, as in cases where there are so many applicants who passed that only high scoring applicants are considered for appointment. That may be the case for the 2002 exam. However, almost universally the least adverse impact is found at the passing point, so if there is adverse impact at the passing point, there certainly will be adverse impact at any higher effective passing point. I therefore did the most conservative analysis.

| Table 16b. Adverse Impact: Pass-Fail Written Test of Exam 2043 | | | | | | |
|--|---------------------------|----------------------------|-------------|------------|--------------------------------|-----------|
| Black: Pass Written | White: Pass Written | Adverse Impact Ratio | Probability | Chi Square | Standard Deviation Units | Shortfall |
| 85.4% | 97.2% | .88 | < .0001 | 476.9 | 21.8 | 150 |

c. Probability and Statistical Significance of Passing Written Test: Exam 2043 (2002)

The adverse impact ratio of .88 is highly statistically significant. A ratio this small would occur by chance much less than one time in ten thousand (see Table 16b, the column titled Probability). This probability corresponds to 21.8 standard deviation units.

d. Practical Import of Passing Written Test: Exam 2043 (2002)

The practical importance of 85.4% of black applicants passing the written test versus 97.2% of white applicants is that there is disproportionate passing of this hurdle, with a shortfall of 150 black passers. Although the adverse impact ratio does not fail the .80 rule of thumb of the *UGESP* the shortfall is of practical significance. Only 79 black applicants²⁴ were appointed as entry level Firefighters as of the last third of 2007.²⁵ Therefore, a shortfall of 150 black passing applicants is of great practical significance.

e. Opinion: Pass-Fail Written Test: Exam 2043 (2002)

There is practically important and highly statistically significant adverse impact against black applicants in passing the written test.

17. Adverse Impact in Passing the Physical Performance Test: Exam 2043 (2002)

In addition to passing the written test, an applicant must pass the Physical Performance Test PPT (with a score of 75) to pass the exam. I did an adverse impact analysis of the pass-fail decisions²⁶ for the PPT of the 2002 exam, as follows.

²⁴Elsewhere I report 101 black applicants were appointed. But there I included as appointed 22 black applicants who were offered appointment but declined (and so were not actually appointed).

²⁵I was not told the effective date of the data file I received. I infer this time period from the date of the last large certification on the data file, 8/2/2007, and the 208 appointments were made from the 498 applicants certified on that date.

²⁶Sometimes the pass-fail point has little practical import, as in cases where there are so many applicants who pass that only high scoring applicants are considered for appointment. That was not the case for the 2002 PPT: at least 2 applicants with the lowest passing score on the PPT were appointed.

shift or other aspects of job assignments, or "acting" assignments (as when a supervisor is on vacation), such a difference in seniority will have career-long detrimental effects on black Firefighters.

e. Opinion on Average Test Score Differences: Exam 2043 (2002)

There is practically important and highly statistically significant adverse impact against black applicants in the average scores on the written test.

22. Adverse Impact in Rank Order Placement On the List: Exam 2043 (2002)

Focusing on the passing applicants, we can ask if there was adverse impact in the ranking of black applicants who passed the exam. This is important because no one can be appointed before they are certified, and certification is in rank order starting with the applicant ranked first on the list (who is the applicant with the highest score). Therefore, I did an adverse impact analysis of the rank order placement, comparing the placement of black and white applicants in two ways: average rank, and proportion in each grouping of 1,000 ranks, as in my analysis of the 1999 exam. My adverse impact analysis of rank order placement on the list follows.³⁰

a. Average Rank on the List: Exam 2043 (2002)

I calculated the average (i.e., the mean) rank on the list for black and white applicants based on an Excel file provided to me titled "7029 Spreadsheet FINAL.xls". The average ranks on the list are presented in Table 22a below.

| Table 22a. Average Rank on the List, by Race | |
|--|------------|
| Group | Mean Rank* |
| Black | 4,598 |
| White | 3,624 |

* Lower numbered ranks are certified before higher numbered ranks.

b. Probability and Statistical Significance of Average Rank on the List: Exam 2043

The difference between the average rank on the list for black and for white applicants is 974 ranks, and this difference is highly statistically significant (see Table 22b). A difference this large would occur by chance less than one time in ten thousand. This corresponds to 9.6 standard deviation units.

³⁰I replicated this analysis for males only, and the results were virtually unchanged as to adverse impact and statistical significance.

| Table 22b. Average Rank on the List, by Race: Statistical Significance | | | | | |
|--|-------|------------|-------------|--------------|--------------------------|
| Black | White | Difference | Probability | F Statistic* | Standard Deviation Units |
| 4,598 | 3,624 | 974 | < .0001 | 91.5 | 9.6 |

* With 1 and 6,523 d.f.

c. Practical Import of Average Rank on the List: Exam 2043 (2002)

The practical importance of black applicants being ranked lower on the list is that, other things being equal, black applicants will be hired later than white applicants. In addition, lower written scores seem to discourage applicants from pursuing the job of Firefighter. The next analysis will show the lower rankings in terms of percentages within groupings of 1,000 ranks.

d. Ranks on the List in Groupings of One Thousand: Exam 2043 (2002)

I divided the list into nominal groupings of 1,000 ranks. Table 22c shows the number and percent of black and white applicants in each grouping of ranks. White applicants are uniformly spread across the groupings of ranks, with some 13 to 14% in each grouping, while black applicants tend to cluster in the bottom groupings of ranks. The top two groupings of 1,000 ranks include 27.3% of the white applicants but only 18.1% of the black applicants. The bottom two groupings include 17.0% of the white applicants but 36.4% of the black applicants, mirroring the totals in the top two ranks, but in reverse. In short, black applicants tend to appear lower on the list.

e. Probability and Statistical Significance of Rankings: Exam 2043 (2002)

The difference just noted in the proportion of black and white applicants between groupings of one thousand ranks is statistically significant (Chi Sq = 138.1, df 7, $p < .0001$). A difference this large would occur by chance less than 1 time in ten thousand.

f. Practical Import of Rankings: Exam 2043 (2002)

The practical importance of black applicants being clustered in the lower rankings is that, other things being equal, black applicants will be hired later than white applicants. To the extent that seniority is a factor in making decisions about shift or other aspects of job assignments, or "acting" assignments (as when a supervisor is on vacation), then a difference in seniority will have career-long detrimental effects on black Firefighters.

g. Opinion on Ranking on the List: Exam 2043 (2002)

Of the applicants who pass the exam, there is practically important and highly statistically significant adverse impact against black applicants in placement on the rank order eligible list.

| Table 22c. Number and Percent in Each Thousand Ranks, by Race | | | |
|--|--|--|-------------------------------------|
| Grouping of Ranks | White: Number in Grouping (percent) | Black: Number in Grouping (percent) | Total of Black and White |
| First 1,000 ranks | 837 (13.8%) | 32 (7.0%) | 869* (100%) |
| Second 1,000 ranks | 820 (13.5%) | 51 (11.1%) | 871* (100%) |
| Third 1,000 ranks | 864 (14.2%) | 42 (9.2%) | 906* (100%) |
| Fourth 1,000 ranks | 847 (14.0%) | 52 (11.3%) | 899* (100%) |
| Fifth 1,000 ranks | 851 (14.0%) | 46 (10.0%) | 897* (100%) |
| Sixth 1,000 ranks | 816 (13.5%) | 69 (15.0%) | 885* (100%) |
| Seventh 1,000 ranks | 752 (12.4%) | 102 (22.2%) | 854* (100%) |
| Eighth 1,000 ranks** | 279 (4.6%) | 65 (14.2%) | 344* (100%) |
| All Ranks | 6,066 | 459 | 6,525 |

* This number is less than 1,000 because there is a total of 960 applicants in other racial or ethnic groups or who did not report an racial or ethnic group.

** There are less than 1,000 ranks in this last grouping.

EXHIBIT S

Report in the matter of
United States of America v. City of New York
and
Vulcan Society Inc. et al. against The City of New York et al.
(CV 07 2067)

Dr. Philip Bobko, Building Personnel Systems LLC
Dr. F. Mark Schemmer, Consultant

January 21, 2008

another), yet the .60-.55=.05 difference in rates might be “statistically significant” in a large city but not statistically significant in a relatively less large city. We note below that New York City’s large size makes it vulnerable to this statistical artifact.

The second sentence in the quoted Guidelines also mentions the notion of practical significance. The concept of practical significance is not well-specified in the academic literature. Both Siskin and Wiesen use alleged “shortfalls” and “delays” as indices of practical significance. We comment on these flawed analyses later.

The Guidelines (in their question and answer section) offer the 80% rule of thumb as an index of practical significance. We note that the Wiesen report does provide results from the 80% rule and that report, like the Guidelines, uses the 80% rule to denote “obvious” practical significance (see, e.g., Wiesen report, p. 16).

Below, we take these concepts in turn (i.e., the 80% rule, then statistical significance, then other forms of practical significance) in addressing the issue of possible adverse impact in exams #7029 and #2043.

II. The 80% rule is useful. Its application demonstrates mixed evidence for any adverse impact in exams #7029 and #2043

As noted above, the 80% rule is the focus of the first sentence in the Guidelines’ discussion of adverse impact (referred to as a rule of thumb in other places in the Guidelines). In our opinion, the Siskin report inappropriately dismisses this rule. The 80% rule is not “arbitrary,” as claimed in the Siskin report.

Rather, the 80% rule has a long history (dating back to the 1970’s) and was eventually used in the Guidelines – a document for which the Department of Justice is one of the four adopting agencies. The Guidelines (question and answer #11) state that the 80% rule “is not intended as a legal definition, but is a practical means of *keeping the attention of the enforcement agencies on serious discrepancies* [italics added for emphasis] in rates of hiring, promotion and other selection decisions.”

As noted, the Wiesen report uses the 80% rule to index obvious practical significance.² Also, academics in personnel selection and adverse impact almost exclusively use the 80% rule (and associated adverse impact ratios) in their research. And, in *USA v. New York City Board of Education et al.* (96-CV-0374, FB, RML), a Memorandum and Order in the Eastern District of New York notes that the “four-fifths” rule is one of the two primary ways the Second Circuit evaluates “*prima facie* disparities under Title VII” (p. 47).

So, what is the result of application of the 80% rule to the NYC firefighter data? We first analyze the Siskin report in this regard (see table below for a summary).

² As noted in his deposition, “a difference that large is a difference worth taking note of” (p. 120).

For example, on pages 21-22 (and Table 1) of the Siskin report, the White pass rate for written exam #7029 is .8992; the Black pass rate is .6026. The ratio of these two rates (i.e., .6026/.8992) is noted as 67.0% (p. 22). In other words, the Black pass rate is not within 80% of the White pass rate.

Now consider pages 23-24 (and Siskin's Table 2) of the Siskin report. The White pass rate for written exam #7029 is .8992 and the Hispanic pass rate is .7667. The ratio of these two rates is reported as 85.3% (see p. 24). In other words, the Hispanic pass rate is within 80% of the White pass rate – and there is no violation of the 80% rule of thumb.

The table below considers all occasions where two pass rates were analyzed in the Siskin report. An application of the 80% rule is then added.³ The result is:

80% rule applied to Siskin report

| <u>analysis</u> | <u>(page # ff.)</u> | <u>ratio of rates</u> | <u>80% standard indicate adverse impact?</u> |
|----------------------------------|---------------------|-----------------------|--|
| 7029 B-W @84.705 | 21 | .670 | yes |
| 7029 H-W @84.705 | 23 | .853 | no |
| 2043 B-W @70 | 26 | .878 | no |
| 2043 H-W @70 | 27 | .955 | no |
| 2043 B-W @eff. | 28 | .590 | yes |
| 2043H-W @eff. | 30 | .838 | no |
| 2043 B-W @ list number eff | 32 | .676 | yes |
| 2043 H-W @ list number eff | 34 | .869 | no |

³ These analyses involve situations where the comparison involves two passing rates and not means or medians, as the 4/5ths rule does not apply in such instances.

Note to above table: all analyses are pass/fail; pass/fail rates were determined at [i.e., @] 70 or 84.705 or at the "effective" (eff) cut.

Note that across the data in both exams, five of the eight ratios do not violate the 80% standard as discussed in the Guidelines. According to this criterion, there is little evidence of "serious" adverse impact in these situations.⁴

A similar, overall summary can be made of the Wiesen report. Thus, the table below delineates any instances of adverse impact ratios (i.e., ratios of rates) for the written and physical components that involved ethnic groups in that report. (That is, some of the analyses in the Wiesen report are specific to males, or are gender-based comparisons. They are not summarized here; nor are the "arbitrarily modified" (p. 59) analyses summarized.)

The summary of the Wiesen report is:

80% rule results in Wiesen report

| <u>analysis</u> | <u>(page # ff.)</u> | <u>ratio of rates</u> | <u>80% rule indicate adverse impact?</u> |
|------------------------------|---------------------|-----------------------|--|
| 7029 appointment | 13 | .28 | yes |
| 7029 written @ 84.705 | 18 | .67 | yes |
| 7029 physical | 19 | .88 | no |
| 7029 overall exam | 24 | .52 | yes |
| 7029 appoint if certified | 30 | .63 | yes |
| 2043 appointment | 38 | .44 | yes |
| 2043 written @ 70 | 42 | .88 | no |
| 2043 physical | 43 | .98 | no |

⁴ In fact, adopting the .80 rule of thumb standard as the null hypothesis in a test of statistical significance, we note that at least four of the adverse impact ratios in the above table are, in fact, statistically significantly greater than .80. (We did not consider a statistical test on the ratio of .869 from the Siskin report, as that report does not provide exact sample size values.)

| | | | |
|------------------------------|----|-----|-----|
| 2043 overall exam | 48 | .76 | yes |
| 2043 appoint if certified | 54 | .89 | no |

NOTE: all analyses in the Wiesen report involve Whites and Blacks; unless otherwise noted (e.g., appointment proportions), analyses are in regard to passing/failing.

Note that four of the ten ratios do not violate the 80% standard as discussed in the Guidelines. Thus, across the two reports, nine of the eighteen ratios do not violate the 80% standard for adverse impact as discussed in the Guidelines.

III. Interpretation of the levels of statistical significance in the Siskin and Wiesen reports are overstated and inconsistent

Some background on statistical significance testing. As noted in the introduction of our report, another way to assess potential adverse impact is to conduct statistical significance tests. Such tests, as outlined in both the Siskin and Wiesen reports, help to decide if any differences in outcomes (across subgroups) are likely due to chance or not. In other words, when assessing two subgroups (e.g., ethnic subgroup differences) a non-zero difference could result, but the non-zero value could be relatively small and due to naturally occurring variation. (A common example is that flipping a fair coin 10 times will not always result in exactly 5 heads and 5 tails.) To repeat, statistical tests are designed to help decide if the obtained difference is likely due to such natural variation (or something else – which is not explained by the statistics test itself).

Research methodologists have generally adopted the following dichotomous decision rule in statistical significance testing: if the difference between two subgroups is less than 5% (or 1%) likely to have occurred by chance, then the difference is said to be “statistically significant.”⁵

Equivalently, methodologists convert the actual difference between two subgroups into the “number of standard deviations” of difference between the two groups. Differences of two or three of these standard deviations correspond (approximately) to the 5% and 1% values above. Hence, as noted in the Siskin (p. 15) and Wiesen (p. 15) reports, social scientists and the courts sometimes refer to these statistical tests as the two (or three) standard deviations analyses. That is, if the subgroup difference is more than two

⁵ Tests of statistical significance can be constructed to compare a variety of things. Given that the focus in this matter is on two subgroups at a time (e.g., Whites and Blacks or Whites and Hispanics), explanations in this section are provided in this two subgroup context for ease of reading.

occur throughout the report (e.g., 6.6 standard deviations is not associated with any adjective [p. 27], yet 5.48 standard deviations is labeled “highly significant” [p. 23].)

[In passing, note that the Siskin report additionally uses the phrase “infinitesimally small” in describing probability levels – which are also a function of sample size. The Siskin report also uses this phrase inconsistently. For example, the phrase appears when linked to a probability as small as 1 in 186,225 yet does not appear when linked to a probability as small as 1 in 2 million (see page 34).]

As for the Wiesen report, it appears to label every statistically significant result as “highly significant” if the number of standard deviations is more than 3.6 in any specific analysis. There is no rationale given for this interpretive stance – and, of course, it is inconsistent with the text in the above footnotes.

Also, in the opening summary of the Wiesen report, the results are described as “very highly significant” (see p. 4) without explanation or direct numerical support for this additional, second adjective.

In sum, the Siskin and Wiesen reports use adjectives to highlight/interpret their findings – but the highlighting leads to potential overstatements. The overstatements occur because one might obtain “highly” significant results by having a large sample size (and not necessarily a large difference). This is particularly important because the earlier use of the 80% rule indicated that the evidence for adverse impact in exams #7029 and #2043 was mixed.

III b. The influence of sample size and the resulting overstatements in the Siskin and Wiesen reports

We now turn to what is an even more important concern with the Siskin and Wiesen reports. As noted several times above, a major concern with tests of statistical significance is their substantial dependence on sample size.

In the current case, the number of applicants is very large (the City of New York is the largest city in America). This means that the statistical tests in these reports are quite sensitive (perhaps overly-sensitive) to any departures from exactly zero difference between subgroups.

This is an important issue – and we devote Appendix A to explain to the reader the role and influence of sample size in statistical significance tests.

To put this another way (and re-visit an earlier comment), suppose there is a small non-zero difference between two subgroups in regard to selection rates. Suppose the same small difference occurs in both New York City and Washington, DC. All else equal, the statistical test is more likely to be significant (and perhaps more likely to be “highly

significant” in the interpretations of Siskin and Wiesen) in New York City, even though the difference is the same in both cities (the 80% rule avoids this artifact).

Another illustration. Consider Dr. Siskin’s analysis of the pass/fail rates for Hispanics for written exam #2043 (see his pages 27 ff. and Table 6). His table indicates that the two pass rates are approximately 97% and 93%. He also states that the difference in these two rates is “statistically significant.” (Although note that these two rates are within 96% of one another – well within the 80% rule.)

As an illustration of the over-sensitive nature of such a statistical test, consider the same total values of applicants in Dr. Siskin’s Table 6 (i.e., 13,878 Whites and 2123 Hispanics), follow the same calculations (i.e., Yates corrected chi-square analysis), and hypothetically assume that 13,496 of the White applicants and 2047 of the Hispanic applicants passed the test. The two selection ratios would now be .972 and .964, respectively. These are very close indeed (i.e., the difference is .008; less than 1%) – and the two rates are within 99% of one another (again, well within the 80% rule).¹² However, the statistical test would indicate this difference is “statistically significant” (i.e., more than two standard deviations). This is over-sensitivity to small differences and does not seem consistent with the Guidelines (which suggest that rates not within 80% of one another might deserve serious attention).¹³

Similar effects of sample size (i.e., effects of over-sensitivity) operate for all statistical significance tests on New York City data conducted in both the Siskin and Wiesen reports. Another illustration (different exam, different subgroup; focus still on pass rates) is offered to help make our point.

Another illustration. Consider the analysis of the pass/fail rates for Blacks and Whites for written exam #7029. This analysis appears on pages 21 ff. (and Table 1) of the Siskin report and pages 18 ff. of the Wiesen report. Both reports indicate the difference is “highly statistically significant.”

Taking the same total values of applicants (i.e., 12,915 Whites and 1749 Blacks), suppose that 11,613 of the Whites and 1545 of the Blacks passed the test. The two selection ratios would now be .899 and .883, respectively. Once again, the rates would be very close (difference is .016, or less than 2%). However, the statistical test would indicate this

¹² In fact, in his deposition statement, Dr. Wiesen described rates that were within 99% of one another as “very close” (p. 101) to absolute parity.

¹³ Our point is even more salient if one assumes – as noted earlier – that Dr. Wiesen labels any result more than 4 standard deviations as “highly” significant. In fact, suppose in these data that 13,494 White applicants pass and 2030 Hispanic applicants pass. Then the two pass rates would be .972 and .956, respectively (i.e., the difference is .016 and the rates are within 98% of one another). However, the statistical test results in more than 4 “standard deviations,” so a difference in rates of .016 would not only be labeled by Dr. Wiesen as “significant” – such a small difference would likely be labeled “highly significant.”

Dr. Siskin's efforts in this regard are somewhat appreciated, although the choice of 90% is unclear. Why 90%? There is no rationale provided for this percentage. Why not 80%? Why not 95%? In fact, in the list of largest American cities (as reported by wikipedia.org) only ten cities have populations that are greater than 10% of the size of New York City's. Cities that have less than 10% of the population of New York City are numerous and include cities such as Washington, DC, Boston, San Francisco, Baltimore, Seattle, and so forth. So, it is not clear if Dr. Siskin's "arbitrary" (see his use of the term, depo, p. 122) reduction in sample size was sufficient.¹⁶

Nonetheless, for the moment, consider the 90% figure used in the Siskin report. Although the Siskin report re-computes results for six analyses, it does not re-compute six other statistical analyses (in fact, all these additional analyses were associated with fewer "standard deviations" than the first six analyses, and the exclusion of these analyses appears to be misleading). So, the re-computations are estimated below (using the same 90% reduction).¹⁷ A 10% sample size assumption was used for comparability to the Siskin report, although many cities in the U.S. have populations less than 10% of New York City's. Hence, this re-computation is conservative. The results are as follows:

Missing re-computations in Siskin report

| <u>analysis begins on</u> | <u>reported # of std. dev.</u> | <u>recomputed # of std. dev.</u> | <u>original claim of statistically significant?</u> | <u>re-computed statistically significant?</u> |
|-------------------------------|------------------------------------|--------------------------------------|---|---|
| p. 24 | 6.48 | 2.05 | yes | yes |
| p. 25 | 4.57 | 1.45 | yes | no |
| p. 31 | 9.45 | 2.99 | yes | yes |
| p. 32 | 9.74 | 3.08 | yes | yes |
| p. 34 | 4.55 | 1.44 | yes | no |
| p. 34 | 5.04 | 1.59 | yes | no |

NOTE: "std. dev." means "standard deviations"

In sum, the missing re-computations indicate that the evidence for statistically significant adverse impact findings is even more mixed than stated in the Siskin report (as with the 80% rule application). The statistical re-computations reflect the type of mixed evidence one might find if exams #7029 or #2043 were used in cities such as Detroit or Indianapolis.

¹⁶ Given the statements about city sizes in this paragraph, Dr. Siskin's statement that "90 percent puts it in the range where you'll see in a lot of other places" (depo, p. 122) is questionable.

¹⁷ The re-computation is straightforward to estimate. A 90% reduction means that the sample size is reduced by a factor of 10. In turn, the number of standard deviations calculated will be reduced by a factor of the square root of ten (i.e., reduced by a factor of about 3.16).

Reduced sample size analysis – Wiesen report. As noted, the Wiesen report appears not to address or consider the influence of the large sample size inherent in the results presented in his report, although when asked about this, he does note “as you get larger and larger numbers of people taking your test, it is more and more likely that you will notice what might be very small differences” (depo, p. 119).

In the summary statement of his report, Dr. Wiesen claims that his significance tests show more extreme values (of probability and standard deviations) than “any other of the many similar tests I recall conducting over the course of my career” (p. 65, fnote. 38). Assuming the veracity of his experiential statement, this extremity could be due to an unusually large sample size in New York. (Dr. Wiesen notes that the second largest sample size he ever dealt with in his career is only “in the range of 1 to 5,000, maybe” [depo. p. 92].)

To consider the influence of sample size, a 90% sample size reduction value was again used to parallel the Siskin report. The reader is also reminded that such a reduction is likely conservative.

Any analyses that reported the “number of standard deviations” were used to re-compute the Wiesen report values. All analyses were, by construction of the Wiesen report, comparisons of Whites to Blacks. The re-computations are as follows:

Re-computations for Wiesen report

| <u>analysis begins on</u> | <u>reported # of std. dev.</u> | <u>recomputed # of std. dev.</u> | <u>original claim of statistically significant?</u> | <u>re-computed statistically significant?</u> |
|-------------------------------|------------------------------------|--------------------------------------|---|---|
| # 7029 starts here | | | | |
| p. 13 | 15.9 | 5.03 | yes | yes |
| p. 18 | 33.9 | 10.72 | yes | yes |
| p. 19 | 4.9 | 1.55 | yes | no |
| p. 24 | 15.8 | 5.00 | yes | yes |
| p. 25 | 39.5 | 12.49 | yes | yes |
| p. 25 | 6.1 | 1.93 | yes | no |
| p. 27 | 6.5 | 2.06 | yes | yes |
| p. 30 | 8.6 | 2.72 | yes | yes |
| p. 34 | 2.1 | 0.66 | yes | no |
| p. 35 | 6 | 1.90 | yes | no |
| #2043 starts here | | | | |
| p. 38 | 8.9 | 2.81 | yes | yes |
| p. 42 | 21.8 | 6.89 | yes | yes |
| p. 43 | 1 | 0.32 | no | no |
| p. 48 | 7.7 | 2.44 | yes | yes |

| | | | | |
|-------|------|-------|------|-----|
| p. 49 | 33.6 | 10.63 | yes | yes |
| p. 49 | 3.6 | 1.14 | yes | no |
| p. 51 | 9.6 | 3.04 | yes | yes |
| p. 54 | 2.4 | 0.76 | yes | no |
| p. 58 | 1.7 | 0.54 | yes* | no |

NOTE: "std. dev." means "standard deviations." Dr. Wiesen's analyses that were related to gender (e.g., analyses using only males) were not included. Nor were the analyses on pages 36-37 included, as these analyses involved a series of sub-analyses at ten hypothetical cuts for written exam #7029. The "*" indicates that Dr. Wiesen used a one-tailed test to claim statistical significance.

In sum, both the 80% rule application and the re-computations of statistical significance (that account for the influence of sample size) indicate that evidence for adverse impact in exams #7029 and #2043 is mixed. The interpretations of statistical "significance" by plaintiffs' experts are exaggerated.

IV. More context and the unrealistic assumptions in the Siskin and Wiesen analyses

As noted in Section III above, the frame of reference for the Siskin/Wiesen tests of statistical significance begins with an assumption of exactly no difference (i.e., 0.00 difference) between two subgroups. All of their standard deviation values (and associated probabilities) are computed based upon comparing what happened in New York City to the value of 0.00 (we remind the reader that plaintiffs' experts speak of "absolute" and "perfect" parity). Given that the job analysis confirmed that a variety of human cognitive abilities were important to being a successful firefighter, the assumption of 0.00 is unrealistic.

Briefly,

- there is an extensive literature in personnel selection indicating that there are non-zero racial and ethnic subgroup differences in tests assessing cognitive abilities
- such research typically notes that standardized Black-White differences are about $d=1.0$ (and other research suggests that within-job, applicant d 's might be around .70 or .80).¹⁸
- even the Wiesen report notes that the resultant patterns of differences "mirrors what is seen nationwide with this type of examination" (p. 37)

¹⁸ The statistic labeled " d " is commonplace in this literature; it is the standardized average test score difference between two subgroups. The numerator of d is the average score for one subgroup minus the average score for the other subgroup; the denominator is the pooled, within-group standard deviation of the scores. In contrast to standard errors or numbers of standard deviation units (see discussion in the Appendix), the magnitude of d is not directly influenced by sample size.

Although this report has focused on the plaintiffs' experts' reports and allegations of adverse impact, we were also asked to comment on issues of job relatedness and business necessity regarding written exams #7029 and #2043. We do so briefly here.

More than 17,000 applicants for New York City firefighter positions were interested enough to sit for written exam #7029; several years later, over 17,000 individuals also sat for written exam #2043. Across these two exams (through August, 2007), the City has hired almost 6000 individuals to fill positions as a firefighter - resulting in both selection ratios being between 15% and 20%. Thus, New York City faced, and continues to face, important issues in identifying the relatively small proportion of applicants from large pools who are best qualified to perform the job of firefighter (and help ensure the public safety of its citizens) in an efficient manner.

Job relatedness. Although the position of firefighter in New York City is not necessarily identical to that in other jurisdictions, it certainly has much in common.³⁴ In addition to literally extinguishing fires, additional elements of the firefighter job include initial and ongoing training in firefighting practices, technical knowledge of equipment use and maintenance, fire prevention, inspection, and public relations. It is also clear that many cognitive demands are placed upon a firefighter. For example, mis-evaluation of priorities at a fire scene, or inappropriately carrying out fire suppression and extinguishing activities, can lead to extraordinary costs (lost lives and/or property damage).

In 1999, Barrett, Polomsky, and McDaniel reviewed and meta-analytically summarized the scientific literature concerning the selection of firefighters using written examinations.³⁵ Across more than 100 samples, they found that cognitive and mechanical comprehension (which included spatial ability) tests were consistent and substantial predictors of both firefighter job performance and training performance.

As a further example of cognitive demands, consider the task cluster of "size-up" in the Morrongiello job analysis for exam #7029 (DCAS report entitled "Firefighter Exam No. 7029: Test Development Report"). The surviving tasks within that cluster include

- determines level of life hazard at scene
- determines appropriate point of entry into structure for assigned position based on location of fire, building type, security concerns, etc.
- determines which tools and equipment should be taken from apparatus based on structure type, location of fire, etc.
- keeps track of position of other firefighters in reference to fire or emergency

³⁴ For example, an earlier firefighter exam was developed by Landy, Jacobs and Associates Inc., a consulting firm in the arena of industrial-organizational psychology. In their work, the Landy et al. analysis empirically invoked validity generalization from results in Washington DC to develop their items for New York City.

³⁵ Barrett, G., Polomsky, M., & McDaniel, M. (1999). Selection tests for firefighters: A comprehensive review and meta-analysis. *Journal of Business and Psychology*, 13, 507-513.

- communicates with civilians at incident scene to obtain necessary information
- evaluates possible building layout, including layout of individual apartments, based on location of fire escapes, and other information
- selects type of extinguisher (e.g., dry chemical, purple-K, foam, water, alcohol, fluoroprotein) needed for a particular situation

The content of these firefighter job tasks (e.g., use of information to determine, communicate, evaluate, and select things) clearly invoke cognitive processes.

The aforementioned DCAS report describes the processes and procedures they followed to update the New York City firefighter job analysis and develop test items for their 1999 written examination. The DCAS job analysis

- updated important task statements from Landy, Jacobs and Associates' prior task list
- collected firefighter importance ratings of tasks and cognitive abilities, as well as links between these two domains (a process that is widely used in industrial-organizational psychology to provide a basis for demonstrating content validity)
- collected the above information using targeted interviews/observations, a focus group, and a job analysis survey completed by a sample of 192 firefighters which included ethnic/racial minority groups and females

The DCAS report also notes that they

- used an ability taxonomy that is largely based on work by Fleishman (the Fleishman taxonomy provides a consistent framework for researchers to examine incumbent and expert perceptions regarding the ability demands of jobs)
- used nine of these abilities to write items for the written exam (nine abilities also formed the basis for items in the earlier Landy, Jacobs and Associates' written exam)
- used and trained panels of incumbent firefighters as item writers and item reviewers, with attention to diversity of these panels. (Using incumbents as preliminary item writers has several potential advantages. It should help ensure that the item content is consistent with the firefighter job. Also, the language usage and reading level of the item text will tend to be consistent with that in the incumbent population.)

In sum, the cognitively-based written exams were developed following standard job analytic and test development procedures – thus speaking to their job relatedness. The reader is reminded that any applicant who passed the written exam was subsequently invited to take the physical exam. Exam scores for the written exam and the physical exam were then standardized, and a composite was formed where each of these two components was weighted 50%.³⁶

³⁶ The physical demands of the firefighter's job, whether in New York City or elsewhere, are perhaps more readily recognized in fire scene activities than the cognitive demands. These activities include substantial physical strength and balance required to hold and direct a charged hose on flames; the stamina to rapidly climb many flights of stairs while dressed in firefighting attire, wearing a breathing apparatus, and carrying heavy equipment; and transporting ladders from fire equipments to the fire scene and erecting them.

Feasibility and business necessity. It is our understanding that assessments of cognitive aptitude are common in civil service public safety positions, including firefighters and law enforcement officers. We suggest that reasons for this usage include:

- the meta-analytic finding that cognitive ability assessments are among the most valid predictors of job performance across many types of jobs and situations³⁷
- the consistent finding that cognitive abilities are important to the firefighter position (e.g., see the above-cited Barrett et al. meta-analysis)
- efficiency and feasibility with respect to the costs and processing of large numbers of applicants
- consistency (which can also be a significant factor in perceptions of fairness or equity) in the methods and processes used to evaluate applicants, and
- various laws, statutes, and regulations governing civil service hiring.

More specifically, as noted above, over 34,000 individuals took written exams #7029 and #2043. The City of New York used an objectively scored, feasibly administered exam to help choose the best qualified applicants (who were then invited to take the physical exam). Particularly given the large numbers of applicants in New York City, the cost of administering and objectively scoring other types of exams at this stage appears to have been prohibitive and not feasible (e.g., structured interviews or work sample examples would be difficult to develop, administer, score, and keep secure and equivalent across all applicants).

It is also the case that New York City Civil Service Law (2007)³⁸ requires the City to provide each firefighter applicant who has participated in the selection process with the opportunity to review his or her responses to questions, as well as the keyed scoring information and the opportunity to file a request for review. That requirement represents a significant impediment to using many other types of selection measures. In particular, allowing for such applicant review makes it difficult, if not impossible, to re-use prior exam questions. In turn, it would be quite difficult to develop parallel or alternative forms of many selection test types (e.g., work samples, structured interviews). In contrast, it can be a relatively more direct process to develop new items for cognitively based abilities. That is, changes to specific content elements (e.g., specific numerical values) within cognitive items can serve as an effective and efficient method for developing relatively parallel forms for future examination cycles.

³⁷ Schmidt, F. & Hunter, J. (1998). The validity of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.

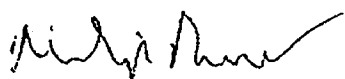
³⁸ Article IV. Recruitment of personnel, Title A. Examinations and eligible lists, Section 50-a, Test validation boards

Jan 21 08 09:40a

Bobko


7173341158

p.2

 1/21/08

Philip Bobko

date

 21-JAN-2008

F. Mark Schermer

date

Appendix A

The role and influence of sample size on the results of statistical significance tests and subsequent numbers of “units of standard deviations”

In statistics, an index of variability in scores is called the standard deviation (symbolized by the letter “s”), which is always a non-negative number. If $s=0$, this means that all scores are equal. If s is bigger than 0, there are some differences in scores. The larger the value of s , the more variation in scores. The value of s is usually calculated across individual scores – i.e., it is an index of how much individual scores vary.

In statistics, there is another type of standard deviation, as well. That is, rather than ask how much variation there is from person to person, one might ask how much variation there is from sample to sample across computed averages (or proportions or passing rates). For example, suppose Researcher A gets a sample and computes an average. Suppose Researchers B and C get different samples and compute averages on their samples. One might then ask, “How much variation is there across the three averages?”

Such a standard deviation (computed as an index of variation across statistics such as averages, or proportions, or passing rates) is sometimes referred to as the “standard error” of the statistic. It is this latter standard error that is used in computing what the courts have come to equate with the phrase “number of standard deviations.”

Thus, the “number of standard deviations” reported in both the Siskin and Wiesen reports actually reflect standard errors (of passing rates, differences between passing rates, averages, etc.). We are not quibbling here with that terminology. Rather, we want the reader to realize that such computations involve standard errors and they are, in turn, substantially influenced by sample size. We try to explain by example. (We use “ n ” to denote sample size; we use a common subgroup sample size to make our point; our largest illustrative n is 2500.)

Suppose the average IQ in the U.S. is 100. In a sample of any size (e.g., $n=400$ is initially chosen for illustration), there is likely to be much variation (i.e., the standard deviation of IQ scores is typically about 15 points and one would expect a range of IQ scores, say, between about 70 and 130 in such a sample).

Now suppose Researcher A computes the average of the IQ scores from this sample of 400. S/he might get 101.4. The number will be fairly close to 100. (It is unlikely all 400 individuals scored near 70 or all scored near 130; there will be a mix.)

Now suppose Researcher B gets a different sample of 400 individuals and computes the average IQ in that second sample. That researcher’s average might be 97.3. Again, there might be much variation within Researcher B’s sample, but it will tend to average out when s/he computes the average.

Similarly, for Researcher C, and so on.

So, within each sample, there might be much variation in scores. However, when one looks at the averages across Researchers A, B, C, etc., those values will all be somewhat close to 100 because, in each instance, use of the average “averages out” the variation.

Of course, the averages calculated by Researchers A, B, C, etc. will not all be exactly equal to 100. They, too, will vary. Hence, one can compute their standard deviation, but note that the variation across these averages will be fairly small.

The statistical nicety is that there is a straightforward relationship between the original standard deviation (s , across individuals) and the standard error (across many researcher’s averages). It is,

$$\text{standard error} = s / (\text{square root of } n)$$

So, in the above example, if the standard deviation (i.e., “ s ”) is 15 IQ points, the standard error of computed averages (based on $n=400$) is $15/(\text{square root of } 400)$, or $15/20$, or 0.75.

As noted, this makes sense. That is, when comparing two individuals, a difference of 15 IQ points would not be unexpected. However, when comparing Researcher A’s average score (across 400 individuals) and Researcher B’s average score (across a different 400 individuals), one would expect those two averages to be much closer (standard error of 0.75).

It is the standard error, and not s directly, that goes into the “number of standard deviation units” reported by Drs. Siskin and Wiesen.

For example, suppose two individual scores differ by 15 points (e.g., scores of 98 and 113). Because s is also 15, one might say the number of standard deviations of difference between 98 and 113 is “one unit of standard deviation.” That is, $(113-98)/(s) = (113-98)/15 = 1.0$.

However, Drs. Wiesen and Siskin are not comparing individual scores – they are comparing aggregations of individual scores (for Black, White, and Hispanic subgroups). Thus, for example, if one subgroup average is 98 and another subgroup average is 113, then the associated/appropriate standard deviation is the standard error (i.e., 0.75). The difference in means (between 98 and 113) would thus be reported as $(113-98)/(0.75) = 20$ “units of standard deviation” – a substantial increase compared to the prior value of 1.0.

Perhaps to see this differently, note the form of the equation in the last sentence:

$$\text{numerator} = \text{difference in averages (or passing rates, etc.)}$$

denominator = standard error = $s/(\text{square root } n)$.

Note that the denominator gets smaller as n gets larger, so the ratio (numerator divided by denominator) gets larger as n gets larger.

If the sample sizes are large - as in New York City - one can see that the calculated "number of standard deviations" will be large, even though the difference remains the same.

For example, using the same score values of 113 and 98 and same value of s (standard deviation), suppose the sample size is increased from $n=400$ to $n=2500$. The standard error would then be equal to $15 / (\text{square root of } 2500) = 15/50 = 0.30$. In turn, the difference between 113 and 98 will now be associated with $(113-98) / .30 = 50$ "units of standard deviation" - although the difference (113-98) remains unchanged.

Sample size influences the calculated "number of standard deviations."